



Assessment of neural network training algorithms for the prediction of polymeric inclusion membranes efficiency

Muhammad Yaqub¹, Beytullah Eren^{2*}, Volkan Eyüpoğlu³

28.04.2016 Geliş/Received, 27.07.2016 Kabul/Accepted

doi: 10.16984/saufenbilder.14165

ABSTRACT

The aim of this study is to introduce, through an appropriate selection of the training algorithm, a better and optimum artificial neural network (ANN) that will capable to predict Polymeric Inclusion Membranes (PIMs) Cr(VI) removal efficiency from aqueous solutions. To accomplish that, three training algorithms including Levenberg-Marquardt (LM), Bayesian Regularization (BR) and Scaled Conjugate Gradient (SCG) have been assessed by training different ANN. The performances of developed models are evaluated by Coefficient of Regression (R^2) and Root Mean Square Error (RMSE) to find the best ANN training algorithms. This study clears that right choice of the training algorithm grants maximizing the predictive capability of the ANN models.

Keywords: neural network, training algorithm, levenberg-marquardt, bayesian regularization, scaled conjugate gradient.

Polimer içerikli membran verimi tahmininde yapay sinir ağları öğrenme algoritmalarının değerlendirilmesi

ÖZ

Bu çalışmanın amacı, polimer içerikli membranlar (PIMs) ile Cr (VI) giderimi için geliştirilecek yapay sinir ağı (YSA) modelinde optimum YSA mimarisi için en uygun öğrenme algoritmasının belirlenmesidir. Bu amaçla, geliştirilen yapay sinir ağı modelinde Levenberg-Marquardt, Bayesian Regularization, Ölçeklenmiş Konjuge Gradyan olmak üzere 3 farklı öğrenme algoritması uygulanmıştır. Ağ mimarisinin ve kullanılan öğrenme algoritmasının ağın tahmin performansına etkisinin belirlenmesinde Regresyon katsayısı (R^2) ve ortalama karesel hata (OKH) teknikleri kullanılmıştır. Sonuç olarak geliştirilen bir YSA modelinde doğru öğrenme algoritması seçiminin ağın tahmin kabiliyeti açısından önemli olduğu sonucuna varılmıştır.

Anahtar Kelimeler: yapay sinir ağları, öğrenme algoritması, levenberg-marquardt, bayesian regularization, ölçeklenmiş konjuge gradyen.

* Corresponding Author

¹Sakarya Üniversitesi, Fen Bilimleri Enstitüsü, Çevre Mühendisliği Anabilimdalı, Sakarya - m_yaqub83@yahoo.com

²Sakarya Üniversitesi, Mühendislik Fakültesi, Çevre Mühendisliği Bölümü, Sakarya - beren@sakarya.edu.tr

³Çankırı Karatekin Üniversitesi, Fen Fakültesi, Kimya Bölümü, Çankırı- volkan@karatekin.edu.tr

1. INTRODUCTION

The contamination of heavy metals in water resources is a very severe problem throughout the world [1], [2]. Chromium is the most hazardous heavy metal and it is extensively used in different industrial application such as metal finishing, leather tanning, electroplating, textile industries, and chromate preparation [3]. Cr(VI) and its compounds are reported as extremely toxic, carcinogenic and mutagenic because they have high water solubility and mobility as well as easy reduction [4].

It is a great challenge to the environmentalists and scientists to develop efficient and economical techniques to remove heavy metals, because, available methods have been found inadequate [5],[6],[7]. Recent research works have shown that Polymeric Inclusion Membranes (PIMs) have potential to remove heavy metals from aqueous solutions i.e. metal ion extraction, separation of inorganic species, biochemical and biomedical applications [8].

To improve an accurate mathematical model to describe all effective parameters is so difficult in PIMs based separation and purification processes. In addition, exact specifying of the separation conditions involved in PIMs process is difficult and may lead to unreliable results in the practical applications. PIMs separation process is complex one that generates non linear data and a trained ANN can duplicate such complicated processes with good accuracy [9] as it has been successfully employed in various environmental engineering applications [10], [11], [12]. In this study, researchers attempted to assess an optimum training algorithm of Artificial Neural Network (ANN) for prediction of PIMs Cr(VI) removal from aqueous solution under various operating parameters.

Literature studies showed that ANNs are prone to either under-fitting or over-fitting [13]. A simple network cannot predict data properly in a complicated data set that causes to under-fitting while on other hand too complex network may fit the noise not just the data, leading to over-fitting which may result in predictions far beyond the experimental data range in training data phase [13]. Therefore, one critical issue in constructing ANN models is to select the best training algorithm for reliable and closely matching results to experimental data.

This study, focused on optimization of ANN modeling performance by changing the training algorithm to find the best architecture with the best predictive capability. The novelty of this work is the application ANN modeling in prediction of PIMs removal efficiency as well as their assessment by using different training

algorithms. It was observed that appropriate choice of training algorithm may have a noteworthy impact on the predictive capability of a network [14].

A performance comparison study of neural network training algorithms in modeling of bimodal drug delivery demonstrated that Levenberg–Marquardt (LM) performed ahead of genetic algorithm [15]. The Bayesian Regularization (BR) outperformed the Cross Validated Early Stopping technique for stream flow forecasting in ANN modeling [16] and also used confidently for modeling of pitting potential [17]. Currently, there are no studies in literature regarding performance comparisons of training algorithms in polymer inclusion membranes Cr(VI) removal efficiency predictions.

Accordingly, the objective of this work is to develop an optimum ANN model for prediction of PIMs Cr(VI) removal efficiency from aqueous solutions by using most suitable training algorithm. To perform that three different ANN training algorithms including Levenberg–Marquardt (LM), Bayesian Regularization (BR) and Scaled Conjugate Gradient (SCG) are employed for prediction of PIMs Cr(VI) removal efficiency. In this work, research efforts are focused on assessing the supervised automated training algorithms to correctly predict the Cr(VI) removal efficiency of PIMs with a minimum error.

2. MATERIALS AND METHODOLOGY

2.1. Experimental Design and Data Preparation

The quantitative Cr(VI) analysis was performed by ICP-MS Agilent 7700x (Santa Clara, USA). To construct an ANN model, precise experimental results are mandatory for training network and testing purpose. In this work, experimental data set comprising of 460 data points was obtained from an experimental research project [18]. The data set of Polymeric Inclusion Membranes (PIMs) operating parameters including time, extractant type, extractant amount, film thickness, plasticizer type and plasticizer amount used as input and removal efficiency as output.

This experimental study regarding the selective transport of Cr(VI) through PVDF-HFP based PIMs containing symmetric imidazolium bromide salts as a carrier. For this reason butyl, hexyl, octyl, and decyl substituted ionic liquids were synthesized and used in the production of PVDF-HFP based PIMs. The polymeric ionic membrane was prepared with using previous methods in the literature [19], [20]. The Cr(VI) transport conditions were optimized with changing PIMs properties (membrane thickness, ionic liquids rate, plasticizer type,

and rate) versus constant aqueous phase (feed and stripping phase) properties, investigated in previous study [21]. The experimental setup and its operation principle were given in Fig. 1.

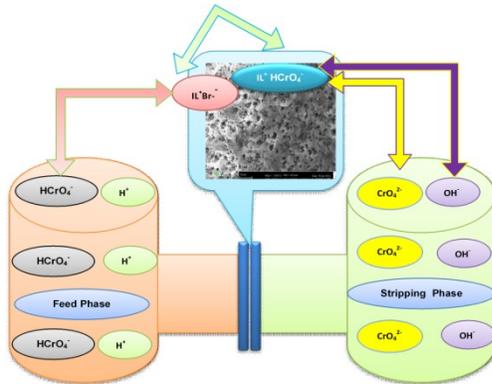


Figure 1. Experimental setup and principle

In this study, the Cr (VI) removal efficiency was predicted by applying three different training algorithms in ANN modeling using available experimental data. MATLAB programming was applied for training and assessment of ANN model with three training algorithms including BR, LM and SCG. Experimental data was randomly divided into three groups for training (70%), validating (15%) and testing (15%) for LM and SCG while testing is (30%) in BR algorithmic technique.

2.2. Artificial Neural Network Description

Artificial Neural Network (ANN) model comprised of an input, a hidden and an output layer. In this study input layer has six nodes as shown in Fig. 2, which conforms to six operating parameters: time, extractant type, extractant amount, film thickness, plasticizer type and plasticizer amount. The output layer contains one neurons that is a dependent parameter representing the PIMs Cr(VI) removal efficiency from aqueous solutions Fig. 2.

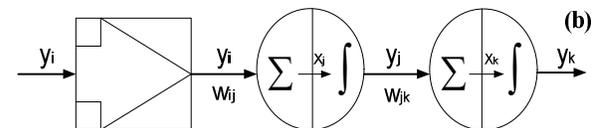
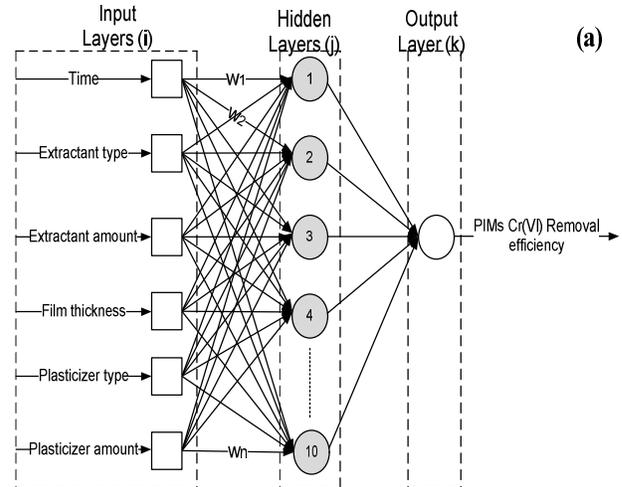


Figure 2. (a) Schematic structure of developed ANN. (b) Function of a neuron

There are six operating parameters in the input layer and the neurons are associated to everyone, according to a chosen network algorithm, by weighted connections through which data signals can travel Fig. 2a. Then all neurons accept numerous inputs (y_i) those are multiplied by their analogous connection weights (w_{ij}) and computed by summation as shown in Fig. 2b [22]. This summation of data is then translated by using a transfer function to outcome a single output (y_j) for that neuron which might possibly moved on to other neurons. Similarly, the predicted results of (y_k) are evaluated by considering the resultant output results of the preceding hidden layer neurons (y_j) as inputs. The appearing signals are multiplied by their respective connection weights and summed up and then approved through the transfer function to generate the resultant output that may be transferred to other neurons [22]. The targeted training algorithms in this study are described in Table 1.

Table 1. Evaluated training algorithms

Training functions	Description
trainlm	Levenberg-Marquardt (LM)
trainbr	Bayesian regularization al (BR)
trainscg	Scaled conjugate gradient (SCG)

2.3. Evaluation of the Training Algorithms

MATLAB programming (R2014a) was applied for ANN modeling and evaluation of three different training algorithms. Since MATLAB applies several learning algorithms [23], [24], three representative training algorithms mentioned in Table.1 have been evaluated: Levenberg-Marquardt (LM), Bayesian Regularization (BR) and Scaled Conjugated Gradient (SCG).

This study describes training and testing of ANN developed models, with three chosen training algorithms for prediction and optimization of ANN models. Initially weights of the ANN models in the training phase were randomly initialized between 0 and 1. The output layer (y_k) was evaluated from the available data set of inputs to the network and the feedback of each neuron and then evaluated results were correlated with the corresponding experimental results.

In the end, prediction error related to the output response was calculated and back circulated to the preceding layers through the network and the weights were regulated to minimize the error by applying three training algorithms as presented in Table 1. The weights were altered once for training data set, the weights were amended 100 times and after that input data set were fed again to ANN model, and were made new predictions. This process was repeated again and again to minimize the error in between experimental and predicted results and then training ended as the error start to rise and that stage is considered as optimized.

In the testing step the capabilities of the trained network was assessed through test data sets, were never exposed aforesaid, were fed to the ANN models. Now, there was no need of corrections for connection weights and the ANN model predicts the PIMs Cr(VI) removal efficiency accurately. The predicted data in the testing step by those three ANNs training algorithms were compared with the real data on the basis of performance criteria in order to choose the most acceptable training algorithm. Evaluated training algorithms are described briefly in following section.

2.3.1. Levenberg Marquardt Algorithm

Levenberg-Marquardt (LM) algorithm is an iterative approach that pinpoint the minimum of a multivariate function that is asserted as the sum of squares of non-linear real-valued functions [25], [26]. Nowadays it has turn into a standard technique for non-linear least-squares problems [27] broadly approved in a wide spectrum of disciplines. It is considered as a merger of steepest descent and the Gauss-Newton method. When the

predicted result is distant from the experimental one, the algorithm performs like a steepest descent method: slow, but assured to converge and if predicted result is close to the experimental, it behaves like a Gauss-Newton method.

2.3.2. Bayesian Regularization

Large weights can cause excessive variance of the output [28] and regularization is a conventional method of handling the negative effect of large weights. The purpose of regularization is to provide the smoother response of network through the adjustment in the objective function by addition of a penalty term that consists of the squares of all network weights. Therefore, small values of weight cause to decrease the propensity of a model to overfit noise in the training data. Bayesian regularization technique was introduced by Mackay [29], which automatically sets the best possible performance function to accomplish the excellent generalization on the basis of Bayesian inference approach. Bayesian optimization of regularization parameters depends upon the calculation of the Hessian matrix at the minimum point [30].

2.3.3. Scaled Conjugate Gradient

Numerous adaptive learning algorithms applicable in feed forward neural networks have newly been introduced [31] and recently number of conjugate gradient algorithms have been introduced as training algorithms in ANN modeling [32], [33], [34]. Several of these algorithms are established on the gradient descent algorithm that is well recognized in optimization theory. Typically they have poor convergence rate and based on parameters which have to be described by the user, because there is no theoretical basis for culling them. The values of these parameters are generally vital for the best performance of the algorithm.

From an optimization point of view training in an ANN model is to minimize the global error function, which is multivariate function that depends upon the weights in the network. This viewpoint provides some positives in the development of a predictive ANN model with effective training algorithm because global error function minimization problem is very common in other fields of science such as conventional numerical analysis [35].

2.4. Data Normalization and Performance Evaluation of ANN Training Algorithm

In this study, before the training of the model both input and output variables were normalized within the range 0 to 1 as follows [36]:

$$x_i = \frac{(x-x_{min})}{(x_{max}-x_{min})} \quad (1)$$

where x_i is the normalized value of a certain parameter, x is the measured value for this parameter, x_{min} and x_{max} are the minimum and maximum values in the database for this parameter, respectively.

The optimized ANN pattern was chosen from different ANN patterns on the basis of their predictive ability performance through statistical techniques such as root mean square error (RMSE) and coefficient of determination R^2 . The RMSE represents the error between model predictions and experimental results. It can be computed with Eq. (2) with a range from 0 to 1. Lower RMSE values are close to zero as preferable as there is no absolute criterion for a “good” value [36].

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_{pred}^i - y_{exp}^i)^2}{n}} \quad (2)$$

In this equation ‘n’ represented the number of target values; and X_{pred}^i and Y_{obs}^i are model predictions and their corresponding experimental values, respectively.

Also, the coefficient of determination, R^2 , linear regression is calculated between the ANN predicted results and the experimental results, used to measure the performance of the network. The R^2 has been estimated through Eq. (3).

$$R^2 = \left[\frac{n \sum_{i=1}^n y_{exp,i} y_{pred,i} - (\sum_{i=1}^n y_{exp,i})(\sum_{i=1}^n y_{pred,i})}{\sqrt{[n \sum_{i=1}^n y_{exp,i}^2 - (\sum_{i=1}^n y_{exp,i})^2] \times [n \sum_{i=1}^n y_{pred,i}^2 - (\sum_{i=1}^n y_{pred,i})^2]}} \right]^2 \quad (3)$$

This value shows the percentage of variability between experimental data and model predictions. R^2 values range between 0 and 1 (i.e. 0–100%) and how much this value has been found close to 1 which means it has a strong positive relationship between predictions and target values [36]. The RMSE and R^2 values provide information on general error ranges between model predictions and experimental results.

3. RESULTS AND DISCUSSION

This study presents the assessment of three different training algorithms of Artificial Neural Networks (ANN) modeling including Levenberg-Marquardt (LM), Bayesian Regularization (BR) and Scaled Conjugate Gradient (SCG). The data series obtained from laboratory batch experiments for prediction of PIMs Cr(VI) removal efficiency and outputs from the developed models are assessed and compared to find the best training algorithm.

3.1. Pre-assessment of Experimental Data

The experimental data used for this study was obtained under different operating conditions such as time; membrane thickness; extractant type and extractant rate; plasticizer type; and plasticizer rate of PVDF-based liquid membranes for Cr(VI) removal. The time (ranges 0-8); extractant type (ranges 1-4); extractant amount (ranges 0-0.343); film thickness of membrane (ranges 41.23-147.83 μ m); plasticizer type (ranges 1-4); and plasticizer amount (ranges 0-0.3377) were used as inputs and removal efficiency of Cr(VI) (ranges 0.13-1.00) was considered as output variable. Statistical information related to the data of each experiment is summarized in Table 2.

Table 2. Data statistics of model variables n=460

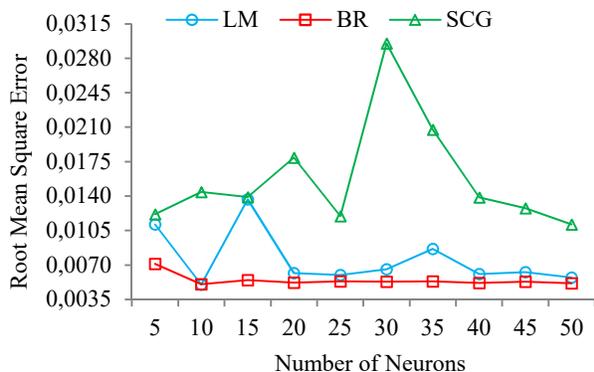
Variables	Data Statistics			
	Xmin	Xmax	Xmean	σ
<i>Input Layer</i>				
Time	0.00	8.00	4.000	2.832
Extractant type	1.00	4.00	2.826	1.167
Extractant amount	0.00	0.34	0.215	0.0679
Film thickness	41.23	147.83	98.953	29.46
Plasticizer type	1.00	4.00	3.673	0.809
Plasticizer amount	0.00	0.34	0.236	0.058
<i>Output Layer</i>				
Removal Efficiency	0.13	1.00	0.865	0.151

X_{min} , X_{max} , X_{mean} : minimum, maximum & mean; σ : standard deviation

3.1. Determination of ANN Topology

The number of layers and number of nodes in each layer have been used to determine the topology of ANN model. The number of neurons (N) in the hidden layer has been determined according to the minimum error prediction which has been considered a basic parameter for ANN structure. In order to determine the optimum number of neurons in the hidden layer, different topologies were examined, in which number of nodes have been found varying from 5 to 50. Each topology has been repeated three times and RMSE has been used as the error function which found 10 neurons in each hidden layer which has been estimated as the best topology due to minimum RMSE for training in LM and BR algorithms. By using

Trial and Error Technique, it was found that (6-10-1) is the best topology in this study as represented in Fig. 3.



3.2. Training Algorithm Selection

The optimum network model was comprised of three layers: an input layer having six operating parameters as inputs, a hidden layer with ten neurons and removal efficiency was considered in output layer. By keeping all parameters constants we can find the better training algorithm. By using (6-10-1) ANN topology as described in above section, three different training algorithms were applied, as summarized in Table 1, and three different ANN models were developed. The topology was kept constant only connection weights were differed on the basis of selected training algorithm. After training, testing set containing of 69 data sets was fed to all developed models to calculate their predictive capability. Statistical tests RMSE and R^2 were used to evaluate if there is any prominent diversity between experimental and predicted results.

3.3. Comparative Study of ANN Training Algorithms

The statistical tests RMSE and R^2 were performed to determine the performance of developed models. Fig. 4 shows the RMSE of three developed ANN models by using results of testing data of three defined training algorithms. In this figure, we presented the test data sets and observed that error minimizes towards end of testing data set. From this analysis it is clear that BR algorithm generates minimum and consistent RMSE as compare to LM and SCG. The SCG is the least performer training algorithm as it gives maximum RMSE and fluctuates widely as shown in Fig. 4.

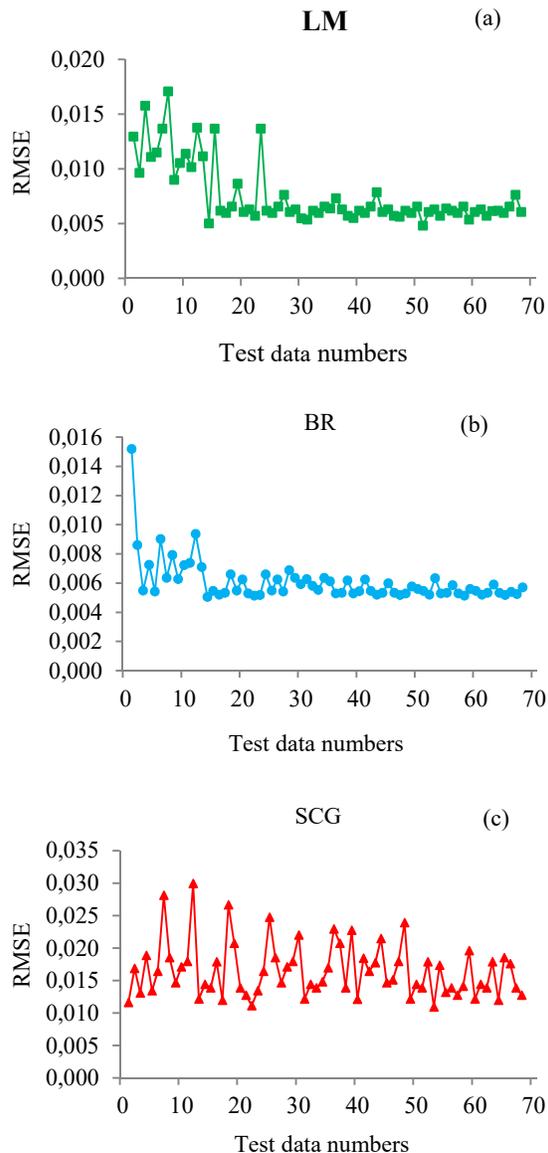


Figure 4. RMSE comparative study of (a) LM, (b) BR and (c) SCG

The Coefficient of Determination (R^2) has been estimated through Eq. (3). This value showed the percentage of variability between experimental data and model predictions. R^2 values range between 0 and 1 (i.e. 0–100%) and how much this value found close to 1 which means it has a strong positive relationship between predictions and target values [36].

The developed ANN models have been assessed by comparing its predicted output results with experimental results through an independent training, validation and testing data sets respectively. The plots of experimental results versus the predicted results have been presented in Fig. 5(a, b, c), 6(a, b, c) and 7(a, b, c) for training,

validation and testing data sets respectively. The results observed as well distributed around X=Y line in a narrow area are considered best results. In training data set results of coefficient of determination were 0.9694, 0.9408 and 0.6958 for LM, BR and SCG training algorithm respectively Fig. 5.1(a, b, c). In validation data set results coefficient values varies as 0.948, 0.9214 and 0.6643 for LM, BR and SCG training algorithm correspondingly Fig. 5.2(a, b, c). The value of $R^2=0.977$ for the line plotted using experimental and predicted testing data set in LM algorithm, and $R^2=0.95$ in case of BR algorithm while $R^2=0.719$ for SCG training algorithm as shown in Fig. 5.3(a, b, c).

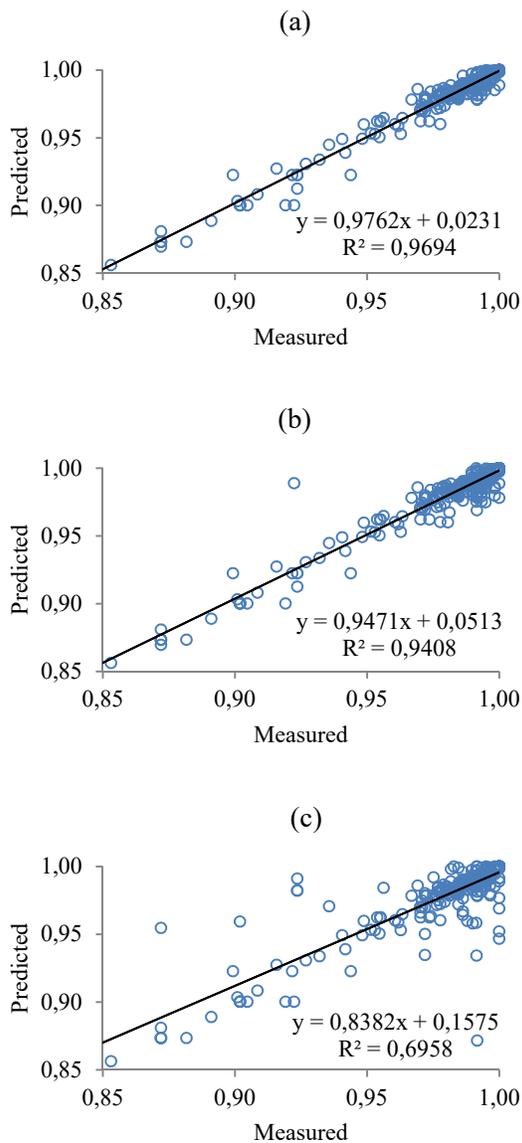


Figure 5. Comparison of experimental and predicted values of training data set by using (a) LM (b) BR (c) SCG
 These results of training, validation and testing data sets demonstrated that LM training algorithm over performed

as compare to BR and SCG while BR algorithm also provided the close results. But results of SCG training algorithm was very poor as compare to LM and BR results as it is shown in Fig. 5.1(a, b, c), 5.2(a, b, c) and 5.3(a, b, c).

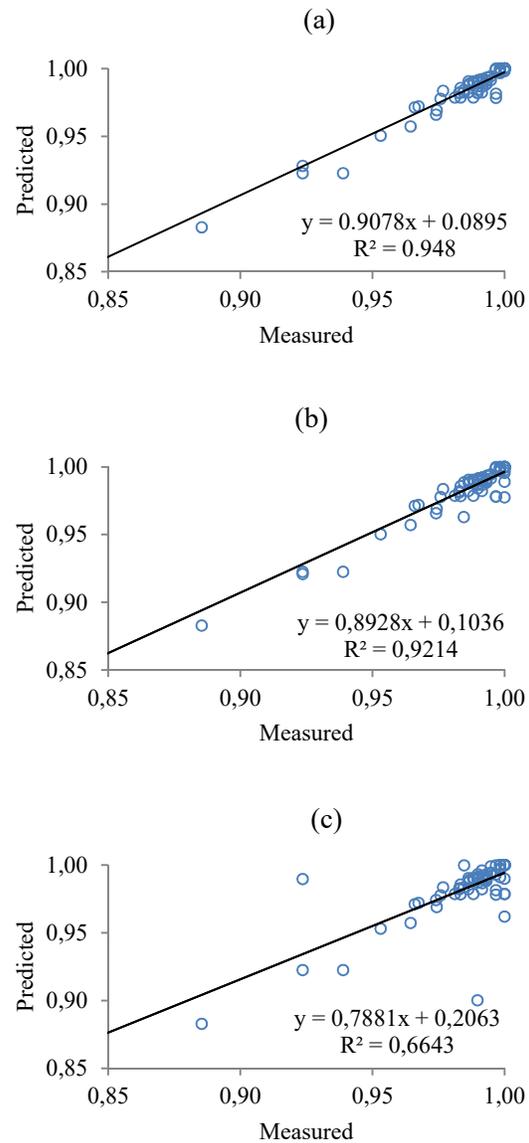


Figure 6. Comparison of experimental and predicted values of validation data set by using (a) LM (b) BR (c) SCG

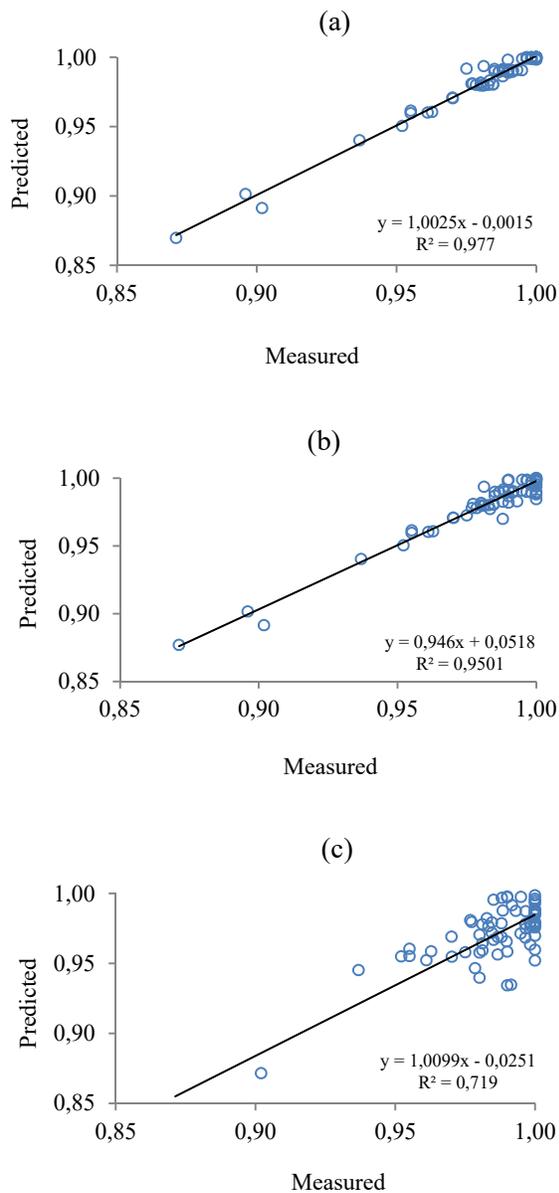


Figure 7. Comparison of experimental and predicted values of testing data set by using (a) LM (b) BR (c) SCG

The comparative study of developed ANN models on the basis of performance criteria such as RMSE and R^2 it is clear that LM training algorithm performed well in training, validation and testing data sets as compare to BR and SCG in prediction of PIMs Cr(VI) removal efficiency.

The comparison of measured, LM, BR and SCG prediction results for the testing data have been graphically presented in Fig. 6. This study clearly depicted that LM training algorithm prediction results

have been found closer to the experimentally measured comparatively better than BR results but best as compare to SCG training algorithm prediction results. The persistent agreement between predicted and experimentally measured results increased the authenticity of the proposed optimum ANN model with LM training algorithm for the prediction of PIMs Cr(VI) removal efficiency. It also has indicated that a well-trained ANN model can be applied to predict Cr(VI) removal efficiency of PIMs without any empirical study which acquire much time, and high experimental costs.

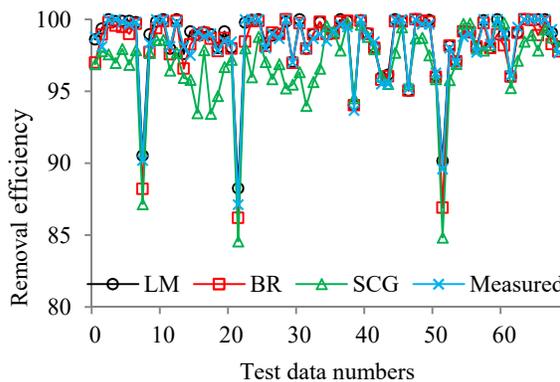


Figure 8. Comparison of the measured, LM, BR and SCG results of testing data set.

4. CONCLUSION

It is concluded that right choice of training algorithm provides maximum prediction capability of ANN models as [14] discussed. Hence, performance of a model is not only dependent on network configuration as mostly expressed in literature, but right selection of training algorithm is also important parameter in optimum ANN model development. In this paper PIMs Cr(VI) removal efficiency was used as example to determine the affect of training algorithms on predictive capability of ANN models.

In this study an ANN model for predictions of Polymeric Inclusion Membranes (PIMs) Cr(VI) removal efficiency has been optimized through a proper selection of the training algorithm. ANN models were developed and trained with three different training algorithms including Levenberg-Marquardt (LM), Bayesian Regularization (BR) and Scaled Conjugate Gradient (SCG), have been evaluated on the basis of their predictive capability. From this computation, LM has been recognized as the best training algorithm for PIMs Cr(VI) removal efficiency predictions. In prediction of PIMs Cr(VI) removal efficiency by using the six inputs including: time, extractant type, extractant amount, film thickness,

plasticizer type and plasticizer amount, LM training algorithm over-performed BR and SCG as indicated by R^2 results as 0.97, 0.95 and 0.72 in testing data for LM, BR and SCG respectively.

Finally, precision of predictive ability was measured for each training algorithm and their performances were in the order of: LM > BR > SCG for training, validation and testing data sets. Moreover, when ANN model is available, approximately infinite combinations of operating parameters within the range of data used in training step can be calculated within short period of time and this provides a significant advantage in designing and optimizing PIMs separation and removal process.

REFERENCES

- [1] B. Volesky and Z. R. Holan, "Biosorption of heavy metals.," *Biotechnol. Prog.*, vol. 11, no. 3, pp. 235–50, 1995.
- [2] F. Veglio' and F. Beolchini, "Removal of metals by biosorption: a review," *Hydrometallurgy*, vol. 44, no. 3, pp. 301–316, 1997.
- [3] Z. Kowalski, "Treatment of chromic tannery wastes," in *Journal of Hazardous Materials*, 1994, vol. 37, no. 1, pp. 137–141.
- [4] V. Gomez and M. P. Callao, "Chromium determination and speciation since 2000," *TrAC - Trends Anal. Chem.*, vol. 25, no. 10, pp. 1006–1015, 2006.
- [5] A. Leusch and B. Volesky, "The influence of film diffusion on cadmium biosorption by marine biomass," *J. Biotechnol.*, vol. 43, no. 1, pp. 1–10, 1995.
- [6] A. Dabrowski, Z. Hubicki, P. Podkoscielny, and E. Robens, "Selective removal of the heavy metal ions from waters and industrial wastewaters by ion-exchange method," *Chemosphere*, vol. 56, no. 2, pp. 91–106, 2004.
- [7] E. S. Z. El-Ashtoukhy, N. K. Amin, and O. Abdelwahab, "Removal of lead (II) and copper (II) from aqueous solution using pomegranate peel as a new adsorbent," *Desalination*, vol. 223, no. 1–3, pp. 162–173, 2008.
- [8] A. K. Pabby, S. S. H. Rizvi, and A. M. Sastre, *Handbook of Membrane Separations Chemical, Pharmaceutical, Food, and Biotechnological Applications*, vol. 1. 2008.
- [9] N. Sipocz, F. A. Tobiesen, and M. Assadi, "The use of Artificial Neural Network models for CO2 capture plants," *Appl. Energy*, vol. 88, no. 7, pp. 2368–2376, 2011.
- [10] Y.-S. Park, T.-S. Chon, I.-S. Kwak, and S. Lek, "Hierarchical community classification and assessment of aquatic ecosystems using artificial neural networks.," *Sci. Total Environ.*, vol. 327, no. 1–3, pp. 105–122, 2004.
- [11] L. Belanche, J. J. Valdes, J. Comas, I. R. Roda, and M. Poch, "Prediction of the bulking phenomenon in wastewater treatment plants," *Artif. Intell. Eng.*, vol. 14, no. 4, pp. 307–317, 2000.
- [12] G. R. Shetty and S. Chellam, "Predicting membrane fouling during municipal drinking water nanofiltration using artificial neural networks," *J. Memb. Sci.*, vol. 217, no. 1–2, pp. 69–86, 2003.
- [13] W. S. Sarle, "Neural Network FAQ, Part 3 of 7: Generalization, periodic posting to the Usenet newsgroup comp. ai. neural-nets," *Retrieved August*, vol. 12, p. 2011, 2002.
- [14] A. P. Plumb, R. C. Rowe, P. York, and M. Brown, "Optimisation of the predictive ability of artificial neural network (ANN) models: A comparison of three ANN programs and four classes of training algorithm," *Eur. J. Pharm. Sci.*, vol. 25, no. 4–5, pp. 395–405, 2005.
- [15] A. Ghaffari, H. Abdollahi, M. R. Khoshayand, I. S. Bozchalooi, A. Dadgar, and M. Rafiee-Tehrani, "Performance comparison of neural network training algorithms in modeling of bimodal drug delivery," *Int. J. Pharm.*, vol. 327, no. 1–2, pp. 126–138, 2006.
- [16] G. & J. . V. WEN WANG, PIETER H.A.J.M. VAN, "Comparing Bayesian Regularization (BR) and Cross Validated Early Stopping for stream flow forecasting with ANN models," *Methodol. Hydrol. (Proceedings Second Int. Symp. Methodol. Hydrol. held Nanjing, China October-November 2005)*, 2007.
- [17] M. J. Jiménez-Come, I. J. Turias, and F. J. Trujillo, "Pitting potential modeling using Bayesian neural networks," *Electrochem. commun.*, vol. 35, pp. 30–33, 2013.
- [18] "Ağır Metallerin Seçici Ekstraksiyonu için İmidazolyum Tuzları İçeren Polimer İçerikli Membranların Üretimi Karakterizasyonu ile Taşıma Verimlerinin Yapay Sinir Ağları ile Modellenmesi, proje no: 112T806, Türkiye Bilimsel ve Teknolojik Araştırma Kurumu," 2015.
- [19] C. Sgarlata, G. Arena, E. Longo, D. Zhang, Y. Yang, and R. A. Bartsch, "Heavy metal separation with polymer inclusion membranes," *J. Memb. Sci.*, vol. 323, no. 2, pp. 444–451, 2008.
- [20] L. M. and M. B. O. Kebiche-Senhadj, "Consideration of Polymer Inclusion Membranes Containing D2EHPA for Toxic Metallic Ion (Pb²⁺) Extraction Recovery," *2015 5th Int. Conf. Environ. Sci. Eng.*, vol. 83, no. 26, p. 30, 2015.
- [21] V. Eyüpoğlu, "Alkyl chain structure-dependent

- separation of Cr(VI) from acidic solutions containing various metal ions using liquid–liquid solvent extraction by butyl-based imidazolium bromide salts,” *Desalin. Water Treat.*, vol. 3994, no. April 2016, pp. 1–16, 2015.
- [22] J. S. Torrecilla, L. Otero, and P. D. Sanz, “Optimization of an artificial neural network for thermal/pressure food processing: Evaluation of training algorithms,” *Comput. Electron. Agric.*, vol. 56, no. 2, pp. 101–110, 2007.
- [23] H. Demuth and M. Beale, “Neural network toolbox for use with MATLAB,” *Citeseer*, 1993.
- [24] V. Vacic, “Summary of the training functions in Matlab’s NN toolbox,” *Matlab*, 2005.
- [25] K. Levenberg, “A method for the solution of certain non-linear problems in least squares,” *Q. Appl. Math.*, vol. 2, pp. 196–168, 1944.
- [26] D. W. Marquardt, “An Algorithm for Least-Squares Estimation of Nonlinear Parameters,” *Journal of the Society for Industrial and Applied Mathematics*, vol. 11, no. 2. pp. 431–441, 1963.
- [27] D. S. Watkins, “The Least Squares Problem,” *Fundam. Matrix Comput.*, no. 1989, pp. 181–259, 2005.
- [28] S. Geman, E. Bienenstock, and R. Doursat, “Neural Networks and the Bias/Variance Dilemma,” *Neural Computation*, vol. 4, no. 1. pp. 1–58, 1992.
- [29] D. J. C. MacKay, “Bayesian Interpolation,” *Neural Comput.*, vol. 4, no. 3, pp. 415–447, 1992.
- [30] D. J. C. MacKay, “A Practical Bayesian Framework for Backpropagation Networks,” *Neural Comput.*, vol. 4, no. 3, pp. 448–472, 1992.
- [31] G. E. Hinton, “Connectionist learning procedures,” *Artif. Intell.*, vol. 40, no. 1–3, pp. 185–234, 1989.
- [32] M. Møller, “A scaled conjugate gradient algorithm for fast supervised learning,” *Neural networks*, vol. 6. pp. 525–533, 1993.
- [33] E. M. Johansson, F. U. Dowla, and D. M. Goodman, “Backpropagation learning for multilayer feed-forward neural networks using the conjugate gradient method,” *Int. J. Neural Syst. J. Neural Syst.*, vol. 2, no. 4, pp. 291–301, 1991.
- [34] K. K. Abbo and H. H. Mohamed, “New Scaled Conjugate Gradient Algorithm for Training Artificial Neural Networks Based on Pure Conjugacy Condition,” vol. 10, no. 3, 2015.
- [35] R. L. Watrous, “Learning Algorithms for Connectionist Networks: Applied Gradient Methods of Non-Linear Optimization,” *Tech. Reports*, no. MS-CIS-87–51, p. 597, 1987.
- [36] E. Dogan, A. Ates, E. C. Yilmaz, and B. Eren, “Application of Artificial Neural Networks to Estimate Wastewater Treatment Plant Inlet Biochemical Oxygen Demand,” *Environ. Prog.*, vol. 27, no. 4, pp. 439–446, 2008.