ESTIMATION OF COAL SWELLING INDEX BASED ON CHEMICAL PROPERTIES OF COAL USING ARTIFICIAL NEURAL NETWORKS

S. Khoshjavan*, M. Heidary and B. Rezai

* *saber.khoshjavan@gmail.com*

Received: January 2010 **Accepted: September 2010**

Department of Mining and Metallurgy Engineering, Amirkabir University of Technology, Tehran, Iran.

Abstract: *Free swelling index (FSI) is an important parameter for cokeability and combustion of coals. In this research, the effects of chemical properties of coals on the coal free swelling index were studied by artificial neural network methods. The artificial neural networks (ANNs) method was used for 200 datasets to estimate the free swelling index value. In this investigation, ten input parameters such as moisture, volatile matter (dry), fixed carbon (dry), ash (dry), total sulfur (organic and pyretic)(dry), (British thermal unit (Btu)/lb) (dry), carbon (dry), hydrogen (dry), nitrogen (dry) as well as oxygen (dry) were used. For selecting the best model for this study the outputs of models were compared. A three-layer ANN was found to be optimum with architecture of ten and four neurons in first and second hidden layer, respectively, and one neuron in output layer. Results of artificial neural network shows that training, testing and validating data's square correlation coefficients (R^2) achieved 0.99, 0.92 and 0.96, respectively. The sensitivity analysis showed that the highest and lowest effects of coal chemical properties on the coal free swelling index were nitrogen (dry) and fixed carbon (dry), respectively.*

Keywords: *Coal Chemical Properties, Free Swelling Index, Artificial Neural Networks (ANNs), Cokeability and Back Propagation Neural Network (BPNN).*

1. INTRODUCTION

Coals are organic sedimentary rocks that have their origin from a variety of plant materials and tissues deposited in more or less aquatic locations. A coal is characterized by a number of chemical, physical, physico-chemical and petrographic properties. In proximate analysis, moisture, ash, volatile matter and fixed carbon are determined. The cokeability of coal is an important technological parameter of coals during the reduction process in the electric furnace method. This is usually determined by the Free Swelling Index [1].

The different models currently available for predicting coke quality take into account selected properties of the coals that make up the blend. Each steel company employs models adapted to its operational characteristics which change as a function of the coals available on the market. However, the role of inerts (breeze, pet coke, anthracite, etc.) added to the coal blend has still not been fully clarified in most of the models. Recently some of the most relevant methods applied in coke quality prediction have been reviewed [2].

The simplest test to evaluate whether a coal is suitable for production of coke is the free swelling index test. It involves heating a small sample of coal in a standardized crucible to around 800 degrees Celsius (1500 °F).

The free swelling index in British Standards Index (BSI) nomenclature (the crucible swelling index number (CSN) in ISO nomenclature) is a measure of increase in the volume of coal when heated, with the exclusion of air. This parameter is useful in evaluating coals for coking and combustion. Coals with a low free swelling index (0-2) are not suitable for coke manufacture. Coals with high swelling numbers (+8) can not be used by themselves to produce coke, as the resultant coke is usually weak and will not support the loads imposed within the blast furnace [3].

When bituminous coals are heated, they develop plastic properties at about 350 °C and as a result. Exhibit fluidity, swelling, expansion and contraction in volume and after carbonization produce a coherent residue the strength of which depends on the ranking of the coal. This plastic property of coals is commonly indicated in the free swelling index, Gieseler plastometry, Ruhr dilatometers, Audibert-Amu dilatometer and Gray-king assay tests.

Gieseler plastometer and Ruhr dilatometer are commonly used to study the plastic properties of coal for coke making. In Gieseler plastometry,

the softening temperature, re-solidification temperature and maximum fluidity of coals are determined to predict their coke ability. In Ruhr dilatometry, the coking capacity, G, defined by Simonis as [4]:

$$G = [(E + V)/2] \times [(c + d)/(V \times c \times E \times)] \quad (1)$$

is used to predict the cokeability of coals.

When the coal particle is heated, its surface becomes plastic while devolatilization occurs from both inside and outside the particle.

Various parameters such as coal type, heating conditions and coal properties affect on free swelling index. For example, Koh Kidena studied the effect of hydrogen/carbon, oxygen/carbon, volatile matter and heating conditions on CSI [5]. In this work the effect of coal chemical properties on swelling index were studied.

After heating for a specified time, or until all volatiles are driven off, a small coke button remains in the crucible. The cross sectional profile of this coke button compared to a set of standardized profiles determines the Free Swelling Index [3].

Artificial neural network (ANN) is an empirical modeling tool, which is analogous to the behavior of biological neural structures [6]. Neural networks are powerful tools that are capable of identifying highly complex underlying relationships from input–output data only [7]. Over the last 10 years, artificial neural networks (ANNs), and, in particular, feedforward artificial neural networks (FANNs) have been extensively studied to present process models, and their use in industry has been rapidly [8].

The aim of the present work is the assessment of properties of more than 200 Illinois coals with reference to the FSI and chemical properties of coal with respect to moisture, volatile matter (dry), fixed carbon (dry), ash (dry), total sulfur (organic and pyretic) (dry), Btu /lb (dry), carbon (dry), hydrogen (dry), nitrogen (dry) and oxygen (dry) using ANNs methods, MATLAB software package.

This work is an attempt to solve the following important questions: (a) which parameters of chemical coal properties have maximum and minimum and positive and negative effect on FSI wide range of Illinois coals? (b) Can we improve the correlation of predicted FSI with actual measured FSI by using of artificial neural network?

2. DATA SET

One of the most important stages in the ANN technique is data collection. The data was divided into training, testing and validating datasets using sorting method to maintain statistical consistency. Datasets for testing and validating were extracted at regular intervals from the sorted database and the remaining datasets were used for training. The same datasets were used for all networks to make a comparable analysis of different architecture. In the present study, 200 datasets were collected among which 25% were chosen for testing and validating. Our data collected from Illinois State Geological Survey website (http://www.isgs.illinois.edu/maps-data-pub/coal-maps/nonconf_masterfile.xls). These data was collected from Illinois state coal mines and geological database.

3. INPUT PARAMETERS

In the present study, input parameters include moisture, ash (dry), volatile matter (dry), fixed carbon (dry), total sulfur (dry), Btu (dry), carbon (dry), hydrogen(dry), nitrogen (dry) and oxygen (dry) to predict the FSI. The ranges of input variables to FSI prediction for the 200 samples are shown in Table 1.

In the present work all inputs (before feeding to the network of matlab software) and output data in training phase, preprocesses the network training set by normalizing the inputs and targets so that they have means of zero and standard deviations of 1:

$$N_p = (A_p - \text{mean } A_{ps}) / \text{std}A_p \quad (2)$$

Where, A_p is actual parameter, mean A_{ps} is mean of actual parameters, $\text{std}A_p$ is standard deviation of actual parameter and N_p is normalized parameter (input). N_p of each data's has been used in network then targets of each parameter converted to actual parameter and after

Table1. The ranges of variables in coal samples (as determined)

Coal chemical properties	Max	Min	Mean	St. Dev.
Moisture (%)	15.94	6.03	10.32	2.21224
Volatile matter, dry (%)	45.10	25.49	36.87	2.458445
Fixed carbon, dry (%)	60.39	30.70	50.58	4.152964
Ash, dry (%)	43.81	4.41	12.56	4.861197
Total sulfur, dry (%)	9.07	0.62	3.00	2.018264
Btu/lb, dry	14 076.00	8 025.00	12 631.08	841.5436
Carbon, Dry (%)	79.32	44.03	70.43	5.026348
Hydrogen, dry (%)	5.36	3.39	4.78	0.310245
Nitrogen, dry (%)	3.03	0.35	1.40	0.290988
Oxygen, dry (%)	12.57	2.16	7.53	1.660288
Free swelling Index	8.50	1.00	4.39	1.268707

these processes figures have been drawn [9].

4. ARTIFICIAL NEURAL NETWORK DESIGN AND DEVELOPMENT INVESTMENT

Neural networks, in general terms, are input–output mapping models that can be used to attack complex or non-straightforward problems. Neural networks are particularly useful in cases where mathematical or statistical methods, such as linear, non-linear regression, curve fitting, etc. cannot provide a satisfactory solution, i.e. the solution can be too general, or too specific that the model cannot react well to new data points. A successful model must have a good generalization and should not be affected by the outlier data points, but at the same time be able to successfully respond when new data points are introduced [10].

Neural networks are simplified models of the biological structure found in human brains [7]. Derived from their biological counterparts, ANNs are based on the concept that a highly interconnected system of simple processing elements (also called “nodes” or “neurons”) can learn complex nonlinear interrelationships existing between input and output variables of a data set [11]. Neural networks are powerful tools that have the ability to identify underlying highly complex relationships from input–output data only [7].

To develop the ANN model of a system, feed-forward architecture namely MLP is most commonly used. This network usually consists of a hierarchical structure of three layers described as input, hidden, and output layers, comprising I, J, and L number of processing nodes, respectively [11]. The general MLP architecture with two hidden layers is shown in Fig 1. When an input pattern is introduced to the neural network, the synaptic weights between the neurons are stimulated and these signals propagate through layers and an output pattern is formed. Depending on how close the formed output pattern is to the expected output pattern, the weights between the layers and the neurons are modified in such a way that next time the same input pattern is introduced, the neural

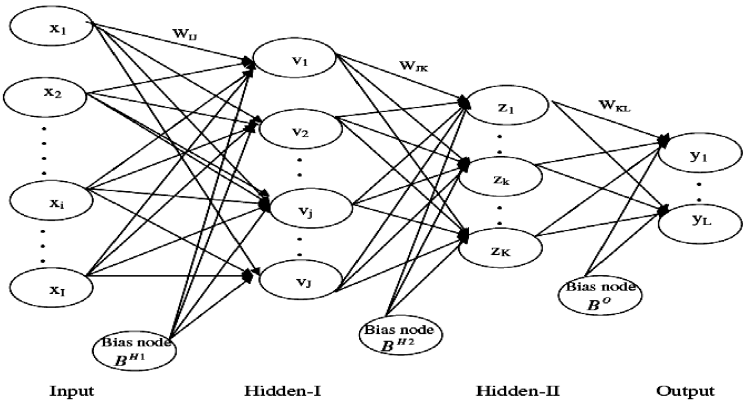

Fig.1. MLP architecture with two hidden layers [12]

network will provide an output pattern that will be closer to the expected response [12].

Various algorithms are available for training neural networks. Feed forward back-propagation algorithm is the most versatile and robust technique, which provides the most efficient learning procedure for multilayer perception (MLP) neural networks. Also, the fact that the back-propagation algorithm is especially capable of solving predictive problems makes it so popular. The network model presented in this article is a supervised back-propagation neural network, making use of the Levenberg-Marquardt approximation.

This algorithm is more powerful than the common used gradient descent methods, because the Levenberg-Marquardt approximation makes training more accurate and faster near minima on the error surface [13].

The method is as follows:

$$\Delta W = \left(J^T J + \mu I\right)^{-1} J^T e \tag{3}$$

In Eq. (3) the adjusted weight matrix ΔW is calculated using a Jacobian matrix J, a transposed Jacobian matrix JT, a constant multiplier m, a unity matrix I and an error vector e. The Jacobian matrix contains the weights derivatives of the errors:

$$J = \begin{bmatrix} \frac{\partial E}{\partial w_{ij}} & . & . \\ . & . & \vdots \\ . & . & \frac{\partial E}{\partial w_m} \end{bmatrix} \tag{4}$$

If the scalar μ is very large, the Levenberg-Marquardt algorithm approximates the normal gradient descent method, while if it is small, the expression transforms into the Gauss-Newton method [7]. For more detailed information the readers are referred to Lines and Treitel [13].

After each successful step (lower errors) the constant m is decreased, forcing the adjusted weight matrix to transform as quickly as possible to the Gauss-Newton solution. When after a step the errors increase the constant m is increased subsequently. The weights of the adjusted weight matrix (Eq. (4)) are used in the forward pass. The mathematics of both the forward and backward pass is briefly explained in the following.

The net input (net_{pj}) of neuron j in a layer L and the output (o_{pj}) of the same neuron of the p th training pair (i.e. the inputs and the corresponding swelling index value of the sample) are calculated by:

$$net_{pj} = \sum_{n=1}^{last} w_{jn} o_{pn} \tag{5}$$

Where, the number of neurons in the previous layer (L -1) is defined by n =1 to last neuron and the weights between the neurons of layer L and L -1 by w_{jn}. The output (o_{pj}) is calculated using the logarithmic sigmoid transfer function:

$$O_{pj} = f_{pj}(net_{pj}) = \frac{1}{1 + e^{-(net_{pj} + \theta_j)}} \tag{6}$$

Where, θ_j is the bias of neuron j.

In general the output vector, containing all o_{pj} of the output layer neurons, is not the same as the

true output vector (i.e. the measured FSI value). This true output vector is composed of the summation of t_{pj}. The error between these vectors is the error made while processing the input-output vector pair and is calculated as follows:

$$E_p = \frac{1}{2}\sum (t_{pj} - o_{pj})^2 \quad (7)$$

When a network is trained with a database containing a substantial amount of input and output vector pairs the total error E_t, (sum of the training errors E_p) can be calculated [7].

$$E_t = \sum E_p \quad (8)$$

To reduce the training error, the connection weights are changed during a completion of a forward and backward pass by adjustments (Δw) of all the connections weights Eqs. (4) and (5) calculate those adjustments. This process will continue until the training error reaches a predefined target threshold error.

Designing network architecture requires more than selecting a certain number of neurons, followed by training only. Especially phenomena such as over fitting and under fitting should be recognized and avoided in order to create a reliable network. Those two aspects - over fitting and under fitting - determine to a large extent the final configuration and training constraints of the network [7].

The number of input and output neurons is the same as the number of input and output variables. For this research, different multilayer network architectures are examined (table 2). Acording to table 2 the best model for network is (10-10-4-1) and the transfer function of model is LOGSIG-LOGSIG-LOGSIG.

During the design and development of the neural network for this study, it was determined that a Four-layer network with 14 neurons in the hidden layers (two layers) would be most appropriate. Artificial neural network (ANN) architecture for predicting of the free swelling index is shown in fig2.

The learning rate of the network was adjusted so that the training time was minimized. During the training, several parameters had to be closely watched. It was important to train the network long enough so it would learn all the examples that were provided. It was also equally important

Table 2. Results of a comparison between some of the models

No	Transfer function	Model	SSE[5]
1	TANSIG-LOGSIG	10-5-1	1.34
2	LOGSIG-LOGSIG	10-7-1	0.7
3	LOGSIG-LOGSIG-LOGSIG	10-4-3-1	1.21
4	TANSIG-TANSIG-LOGSIG	10-5-3-1	1.02
5	LOGSIG-LOGSIG-LOGSIG	10-6-4-1	0.46
6	LOGSIG-LOGSIG-LOGSIG	10-7-4-1	0.3
7	LOGSIG-LOGSIG-LOGSIG	10-8-4-1	0.18
8	LOGSIG-LOGSIG-LOGSIG	10-8-6-1	0.03
9	LOGSIG-LOGSIG-LOGSIG	10-10-4-1	0.014

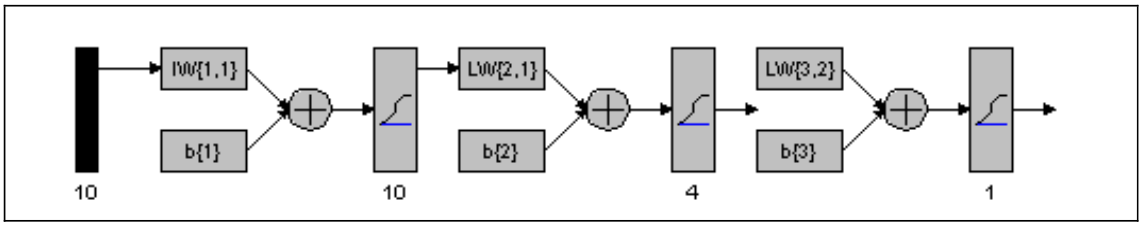

Fig.2. ANN architecture for predict the Free Swelling Index

to avoid over training, which would cause memorization of the input data by the network. During the course of training, the network is continuously trying to correct itself and achieve the lowest possible error (global minimum) for every example to which it is exposed. The network performance during the training process is shown in Fig.3, where the optimum epochs of train achieved 400 epochs.

To evaluate a model, a comparison between the predicted and measured values of FSI can be fulfilled. To this end, MAE (E_a) and mean relative error (E_r) can be used. Ea and E_r are computed as follows [7].

$$E_a = \left| T_i - O_i \right| \tag{9}$$

$$E_r = \left(\frac{\left| T_i - O_i \right|}{T_i} \right) \tag{10}$$

Where T_i, O_i and represent the measured and predicted output.

For the optimum model E_a and E_r were equal to 0.02627 and 0.006635 respectively. Comparison between measured and predicted free swelling index for training, testing and validating datas are shown in Figs. 4, 5 and 6 respectively. Correlations achieved from these figures, between measured and predicted free swelling index from training, testing and validating data, indicate that the network has high ability for predicting free swelling index (Figs.7, 8 and 9).

5. SENSITIVITY ANALYSIS

A useful concept has been proposed to identify the significance of each factor (input) on the factors (outputs) using a trained network. This enables us to hierarchically recognize the most sensitive factors affecting the coal swelling

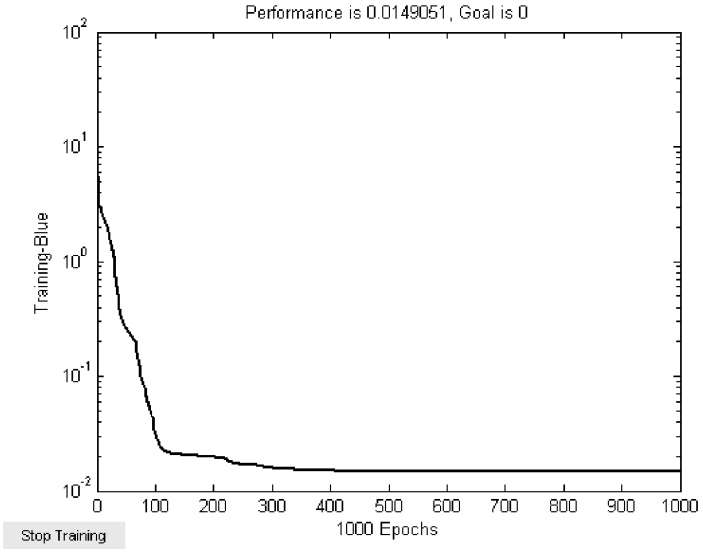

Fig.3. Network performance during the training process

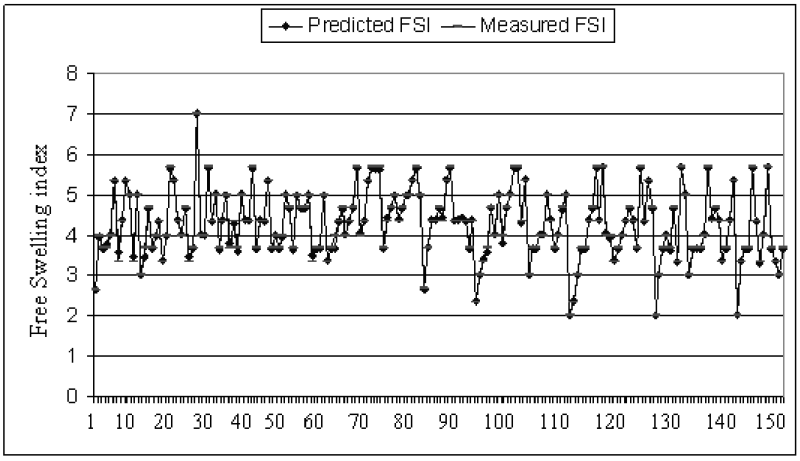

Fig.4. Comparison of measured and predicted Free Swelling Index for different samples of training data

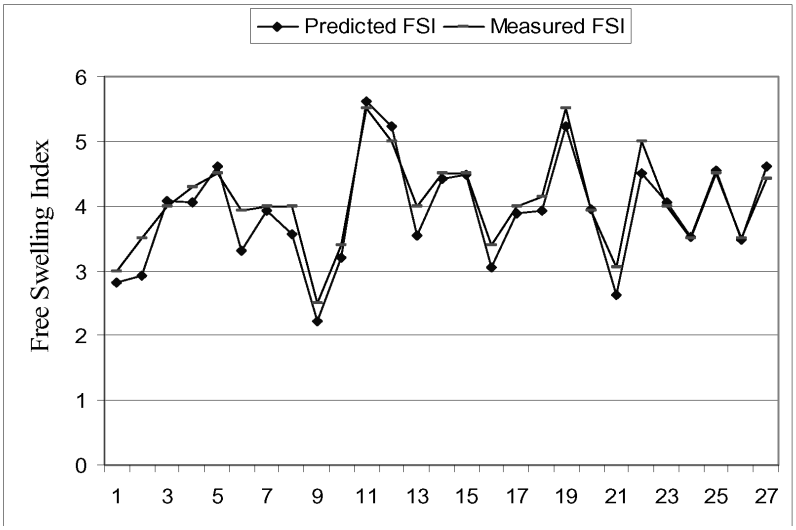

Fig.5. Comparison of measured and predicted Free Swelling Index for different samples for testing data

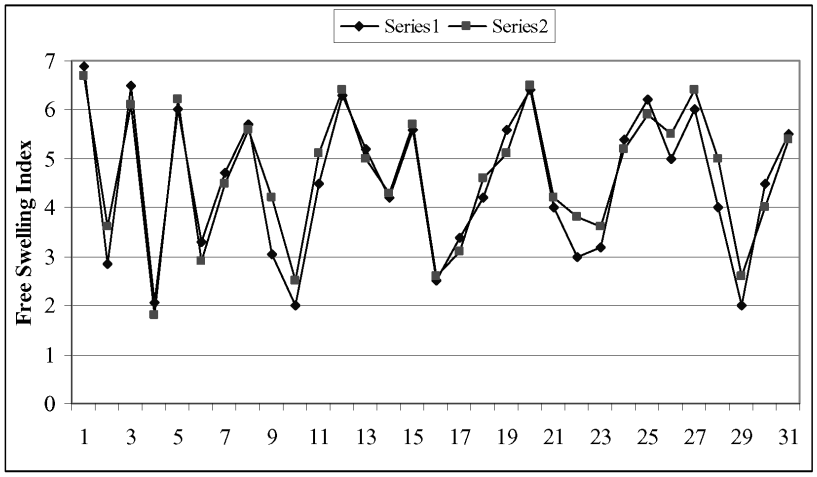

Fig.6. Correlation between measured and predicted Free Swelling Index for validating data

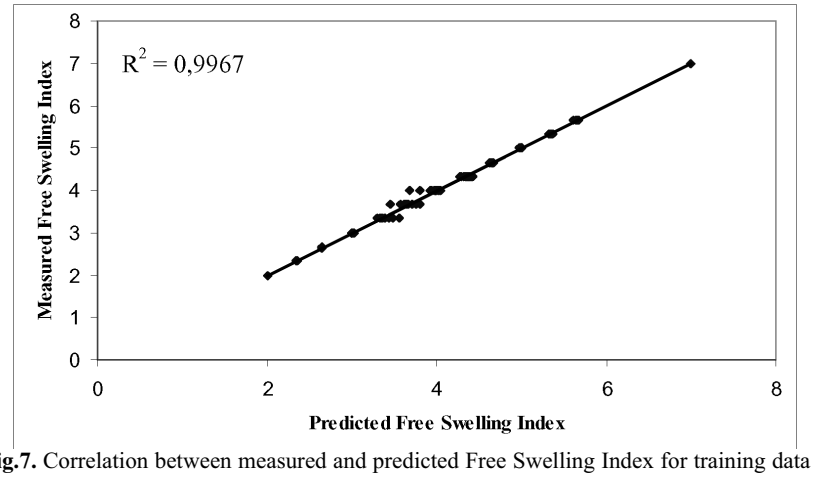

Fig.7. Correlation between measured and predicted Free Swelling Index for training data

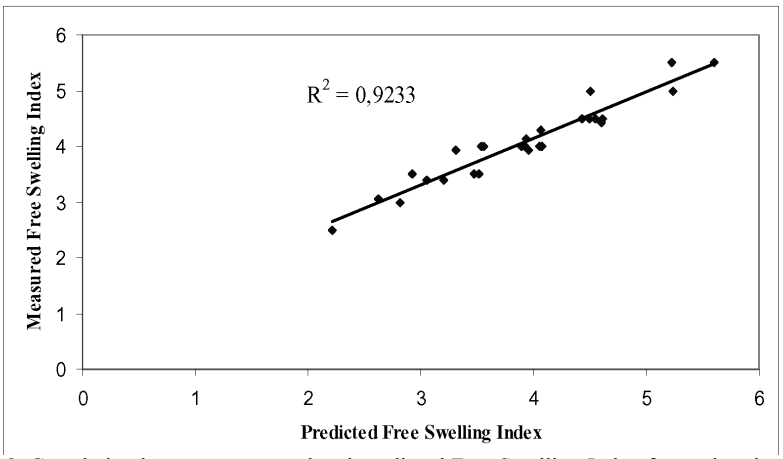

Fig.8. Correlation between measured and predicted Free Swelling Index for testing data

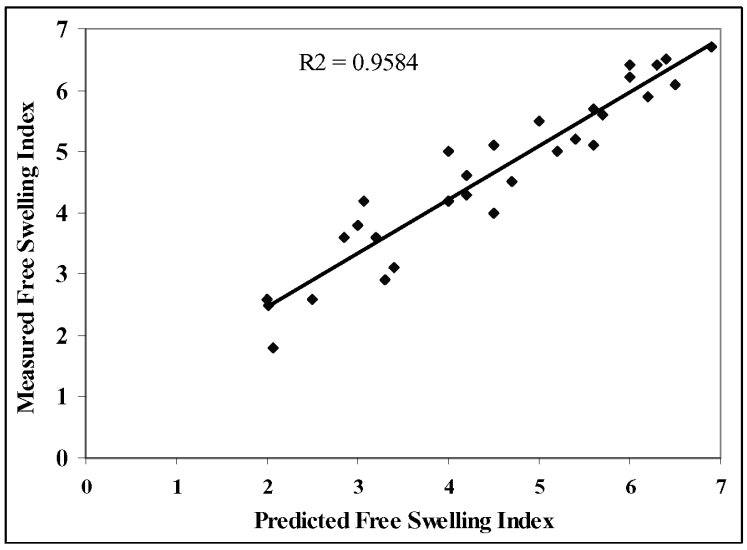

Fig.9. Correlation between measured and predicted Free Swelling Index for validating data

index. This is performed by incorporating values of 'relative strength of effect' (RSEs) [14]. After a BPNN has been trained successfully, the neural network is no longer allowed to adapt. The output for a one-hidden-layer network can be written as:

$$o_k = \frac{1}{(1+e^{-e_k})} \tag{11}$$

Where

$$e_k = \sum_j o_j w_{jk} + \theta_k, \qquad o_j = 1/(1+e^{-e_j}) \tag{12}$$

$$e_j = \sum_j o_i w_{ij} + \theta_j \tag{13}$$

Where, w is a connected weight, Θ is a threshold and o_i is the value of input unit. Thus, we have.

$$O_k = 1 / \left(1 + e^{\left(-\left(\sum_j w_{jk} \left(1 / \left(1 + e^{-\left(\sum_i w_{ij} o_i + \Theta_j \right)} \right) \right) \right) + \Theta \right)} \right) \tag{14}$$

Since the activation function is sigmoid Eq.(11), it is differentiable. The variance of O_k with the change of O_j for a network with n hidden layers can be calculated by the differentiation of the following equation:

$$\partial O_k / \partial O_i = \sum_{jn} \sum_{jn-1} \cdots \sum_{j1} w_{j_n k} G(e_k) w j_{n-1} j_n G(e_{j_n}) \times w_{j_{n-2} j_{n-1}} G(e_{j_{n-1}}) \ldots w_{ij1} G(e_{j1}) \tag{15}$$

Where $G(e_k) = e^{-ek}/(1+e^{-ek})^2$, O_{jn}, O_{jn-1}, O_{jn-2}, … O_{j1} denote the hidden units in the n, n-1, n-2, …, 1 hidden layers, respectively [14]. Obviously, no matter what the neural network approximates, all items on the right hand side of Eq.(15) always exist. According to Eq. (15), a new parameter RSE_{ki} can be defined as the RSE for input unit i on output unit k [14].

Definition of RSE: For a given sample set $S=\{s_1, s_2, s_3, \ldots, s_j, \ldots, s_r\}$ where $S_j=\{X,Y\}$, $X=\{x_1,x_2,x_3, \ldots, x_p\}$, $Y=\{y_1, y_2, y_3, \ldots, y_p\}$, if there is a neural network trained by back-propagation algorithm with this set of samples, the RSE_{ki} exists as:

$$RSE_{ki} = C \sum_{jn} \sum_{jn-1} \cdots \sum_{j1} w_{j_n k} G(e_k) w j_{n-1} j_n G(e_{j_n}) \times w_{j_{n-2} j_{n-1}} G(e_{j_{n-1}}) \ldots w_{ij1} G(e_{j1}) \tag{16}$$

Where C is a normalized constant which controls the maximum absolute value of RSE_{ki} as unit and the function G denotes the differentiation of the activation function. G, w and e are all the same as in Eq. (15).

It should be noted that the control of RSE is done with respect to the corresponding output unit, which means all RSE values for every input unit on corresponding output unit are scaled with the same scale coefficient. Hence, it is clear that RSE ranges from -1 to 1 [14].

Compared with Eq. (15), RSE is similar to the derivative except for its scaling value. But it is a different concept from the differentiation of the original mapping function. RSE is a kind of parameter which could be used to measure the relative importance of input factors to output units, and shows only the relative dominance rather than the differentiation of one to one input and output. The larger the absolute value of RSE, the greater the effect the corresponding input unit on the output unit will be. Also, the sign of RSE indicates the direction of influence, which means a positive action applies to the output when RSE>0, and a negative action applies when RSE<0. Here, a positive action denotes that the output increases with the increment of the corresponding input, and decreases with the reduction of the corresponding input. On the contrary, negative action indicates that the output decreases when the corresponding input increases and increases when the corresponding input decreases. The output has no relation with the input if RSE=0. RSE is a dynamic parameter which changes with the variance of input factors. In a further section, the RSE will be used for a sensitivity analysis of the influence of factors on the free swelling index predicted by a trained neural network.

Fig.8 shows the average RSE values of the factors calculated for all 200 field data that are

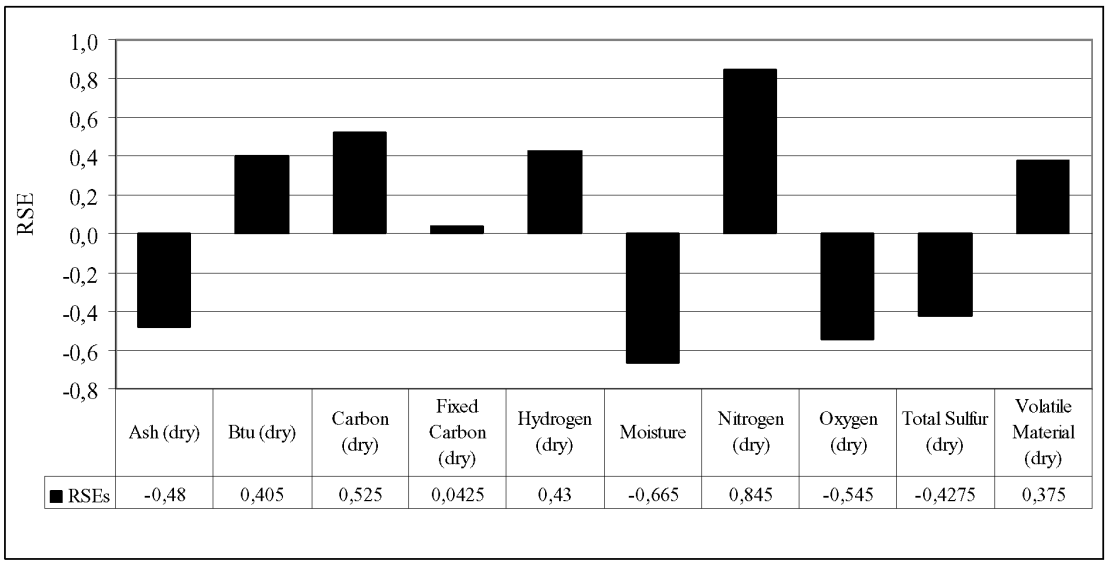

Fig.10. Sensitivity analysis between the free swelling index and coal chemical properties

used in the previous sections. It can be seen in Fig. 8 that 'moisture' and 'nitrogen' are usually the most sensitive parameters. Remaining factors including Btu (dry), carbon (dry), fixed carbon (dry), hydrogen (dry), oxygen (dry), total sulfur (dry) and volatile matter (dry) were also studied by the neural network method. In addition, a positive value of RSE indicates that, for example, if 'carbon' has a positive RSE (see Fig. 10) increases the value of RSE, the FSI will increases, and inverse effects will take place in the case of negative RSE (i.e. 'ash', etc.)..

6. CONCLUSIONS

The authors have drawn the following conclusions:

1. The optimum ANN architecture has been found to be ten and four neurons in the first and second hidden layer, respectively, and one neuron in the output layer.
2. In the ANNs method, results of the artificial neural network show that square correlation coefficients of the training, testing and validating data (R^2) achieved 0.9967, 0.9181 and 0.9584 respectively.
3. Network RSEs show that nitrogen (dry) (+0.845), moisture (-0.665), oxygen (dry) (-0.545), carbon (dry) (+0.525), ash (dry) (-0.48), total sulfur (dry) (-04275), Btu (dry) (+0.425) and volatile matter (dry) (+0.375) were parameter effective on the free swelling index (Fig.10).
4. For network training performance, when the number of epochs is 400, error of training network was minimized and after this point suitable performance was achieved.
5. RSE of Fixed carbon (dry) was +0.0425 and it had lowest effect on free swelling index.
6. Results from the neural network showed that nitrogen (dry), carbon (dry), hydrogen (dry), Btu (dry), volatile matter (dry) and fixed carbon (dry) had positive effects on the free swelling index. With increasing these parameters free swelling index of coal will be increased. The negative effects of input parameters were related to moisture, oxygen (dry), ash (dry) and total sulfur (dry). These parameters had reverse effects on the free swelling index.

REFERENCES

1. Loison, R., Foch, P., and Boyer, A., 1989, Coke quality and production, 2nd edition, Butterworths, London.
2. R. Alvarez, M.A. Dý´ez, and C. Barriocanal, 2007, An approach to blast furnace coke quality prediction, Fuel 86, 2159–2166.
3. Larry T., 2002, Coal Geology, Wiley Publisher.

4. Price, J. T. and Grandsen, J. F., 1987, Improving coke quality with Canadian coals. In Proc, 1st International coke making congress, Essen, section E3.
5. Kidena K., 2007, Prediction of thermal swelling behavior on rapid heating using basic analytical data, Energy & Fuels, 21, 1038-1041. http://pubs.acs.org/
6. Yao H. M., Vuthaluru H. B., Tade M. O., Djukanovic D., 2005, Artificial neural network-based prediction of hydrogen content of coal in power station boilers, Fuel, 84, 1535–1542.
7. Haykin S., 1999, Neural networks, a comprehensive foundation, USA, 2ed. USA, Prentice Hall.
8. Ungar L. H., Hartman E. J., Keeler J. D., Martin G. D., 1996, Process modeling and control using neural networks, American Institute of Chemical Engineers Symposium Series, 92, 57–66.
9. Demuth, H.; and Beale, M., 2002, Neural Network Toolbox for use with MATLAB, Handbook, 154.
10. Özbayoðlu Gülhan, Özbayoðlu A., Özbayoðlu M., 2008, Estimation of Hard grove grind ability index of Turkish coals by neural networks, Int. J. Miner. Process 85, 93-100.
11. Vuthaluru H. B., Brooke R. J., Zhang D. K., Yan H. M., 2003, Effects of moisture and coal blending on Hardgrove Grind ability Index of Western Australian coal, Fuel Processing Technology, 67-76.
12. Shagufta U. Patel. Et al, 2007, Estimation of gross calorific value of coals using artificial neural networks, Fuel, 86, 334–344.
13. Lines LR, Treitel S., 1984, A review of least-squares inversion and its application to geophysical problems, geophysical prospecting, 32.
14. Kim C. Y., Bae G. J., Et al, 2001, Neural network based prediction of ground surface settlements due to tunneling, Computers and Geotechnics 28, 447-517.