



# Evaluation of the effect of coal chemical properties on the Hardgrove Grindability Index (HGI) of coal using artificial neural networks

by S. Khoshjavan\*, R. Khoshjavan†, and B. Rezai\*

## Synopsis

In this investigation, the effects of different coal chemical properties were studied to estimate the coal Hardgrove Grindability Index (HGI) values index. An artificial neural network (ANN) method for 300 data-sets was used for evaluating the HGI values. Ten input parameters were used, and the outputs of the models were compared in order to select the best model for this study. A three-layer ANN was found to be optimum with architecture of five neurons in each of the first and second hidden layers, and one neuron in the output layer. The correlation coefficients ( $R^2$ ) for the training and test data were 0.962 and 0.82 respectively. Sensitivity analysis showed that volatile material, carbon, hydrogen, Btu, nitrogen, and fixed carbon (all on a dry basis) have the greatest effect on HGI, and moisture, oxygen (dry), ash (dry), and total sulphur (dry) the least effect.

## Keywords

coal chemical properties, Hardgrove Grindability Index, artificial neural network, back-propagation neural network.

## Introduction

Coal is a heterogeneous substance that consists of combustible (organic matter) and non-combustible (moisture and mineral matter) materials. Coal grindability, usually measured by the Hardgrove Grindability Index (HGI), is of great interest since it is an important practical and economic property for coal handling and utilization, particularly for pulverized-coal-fired boilers.

The grindability index of coal is an important technological parameter for understanding the behaviour and assessing the relative hardness of coals of varying rank and grade during comminution, as well as their coke-making properties. Grindability of coal, which is a measure of its resistance to crushing and grinding, is related to its physical properties, and chemical and petrographical composition (Özbayoğlu, Özbayoğlu, and Özbayoğlu, 2008). The investigation of the grindability of coal is important for any kind of utilization such as coal beneficiation, carbonization, and many others. The energy cost of grinding is significant at 5 to 15 kWh/t (Lytle, Choi, and Prisbrey, 1992).

Ural and Akyildiz (2004), Jorjani *et al.*, (2008), and Vuthaluru *et al.* (2003) studied the effects of mineral matter content and elemental analysis of coal on the HGI of Turkish, Kentucky, and Australian coals respectively. They found that water, moisture, coal blending, acid-soluble mineral matter content, and  $Na_2O$ ,  $Fe_2O_3$ ,  $Al_2O_3$ ,  $SO_3$ ,  $K_2O$ , and  $SiO_2$  contents affect the grindability of coals. High ash and water- and acid-soluble content samples were found to have higher HGI values, whereas samples with high ash, high  $TiO_2$ ,  $MgO$ , and low water- and acid-soluble contents had lower HGI values. The relationships between grindability, mechanical properties, and cuttability of coal have been investigated by many researchers, who established close correlations between HGI and some coal properties. Tiryaki (2005) showed that there are strong relationships between the HGI of coal and its hardness characteristics.

In order to determine the comminution behaviour of coal, it is necessary to use tests based on size reduction. One of common methods for determining the grindability of coal is the HGI method. Soft or hard coals were evaluated for the grindability - toward 100. A coal's HGI depends on the coalification, moisture, volatile matter (dry), fixed carbon (dry), ash (dry), total sulphur (organic and pyritic, dry), Btu/lb (dry), carbon (dry), hydrogen, nitrogen (dry), and oxygen (dry) parameters. For example, if the carbon content is more than 60%, the HGI moves to maximum range (Parasher, 1987).

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

† Department of Industrial Engineering, Faculty of Engineering, Torbat Heydariyeh Integrating Higher Education.

© The Southern African Institute of Mining and Metallurgy, 2013. ISSN 2225-6253. Paper received Jun. 2009; revised paper received Feb. 2013.

## Evaluation of the effect of coal chemical properties on the Hardgrove Grindability Index

An artificial neural network (ANN) is an empirical modelling tool that behaves in a way analogous to biological neural structures (Yao *et al.*, 2005). Neural networks are powerful tools that have the ability to identify underlying highly complex relationships from input-output data only (Haykin, 1999). Over the last 10 years, ANNs, and in particular feed-forward artificial neural networks (FANNs), have been extensively studied to develop process models, and their use in industry has been growing rapidly (Ungar *et al.*, 1996). In this investigation, ten input parameters – moisture, volatile matter (dry), fixed carbon (dry), ash (dry), total sulphur (organic and pyritic, dry), Btu/lb (dry), carbon (dry), hydrogen (dry) nitrogen (dry), and oxygen (dry) – were used.

In the procedure of ANN modelling the following steps are usually used:

1. Choosing the parameters of the ANN
2. Data collection
3. Pre-processing of database
4. Training of the ANN
5. Simulation and modelling using the trained ANN.

In this paper, these stages were used in the developing of the model.

### Material and methods

#### Data-set

The collected data was divided into training and testing data-sets using sorting method to maintain statistical consistency. Data-sets for testing were extracted at regular intervals from the sorted database and the remaining data-sets were used for training. The same data-sets were used for all networks to make a comparable analysis of different architectures. In the present study, more than 300 data-sets were collected, among which 10% were chosen for testing. This data was collected from Illinois state coal mines and geological database ([www.igs.illinois.edu](http://www.igs.illinois.edu)).

#### Input parameters

The Input parameters for evaluating the HGI comprised moisture, ash (dry), volatile matter (dry), fixed carbon (dry), total sulphur (dry), Btu (dry), carbon (dry), hydrogen(dry), nitrogen (dry), and oxygen (dry). The ranges of input variables to HGI evaluation for the 300 samples are shown in Table I.

#### Artificial neural network design and development

ANN models have been studied for two decades, with the objective of achieving human-like performance in many fields of knowledge engineering. Neural networks are powerful tools that have the ability to identify underlying highly complex relationships from input-output data only (Plippman, 1987; Khoshjavan, Rezai, and Heidary, 2011). The study of neural networks is an attempt to understand the functionality of the brain. Essentially, ANN is an approach to artificial intelligence, in which a network of processing elements is designed. Further, mathematical methods carry out information processing for problems whose solutions require knowledge that is difficult to describe (Khoshjavan, Rezai, and Heidary, 2011; Zeidenberg, 1990).

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 (Tiryaki, 2005).

For developing an ANN model of a system, feed-forward architecture, namely multiple layer perception (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 processing nodes respectively (Tiryaki, 2005). A general MLP architecture with two hidden layers is shown in Figure 1. When an input pattern is introduced to the neural network, the synaptic weights between the neurons are stimulated and these signals propagate through the 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 network will provide an output pattern that will be closer to the expected response (Patel *et al.*, 2007).

Various algorithms are available for training of neural networks. The feed-forward back-propagation algorithm is the most versatile and robust technique, which provides the most efficient learning procedure for MLP neural networks. Also, the fact that the back-propagation algorithm is particularly capable of solving predictive problems makes it so popular. The network model presented in this article was developed in Matlab 7.1 using a neural network toolbox, and is a supervised back-propagation neural network making use of the Levenberg-Marquardt approximation.

Table I  
The ranges of variables in coal samples (as determined)

Coal chemical properties	Max.	Min.	Mean	St. Dev.
Moisture (%)	15.94	6.03	10.32	2.212 24
Volatile matter (dry) (%)	45.10	25.49	36.87	2.458 445
Fixed carbon (dry) (%)	60.39	30.70	50.58	4.152 964
Ash (dry) (%)	43.81	4.41	12.56	4.861 197
Total sulphur (organic and pyritic) (dry) (%)	9.07	0.62	3.00	2.018 264
Btu/lb (dry)	14 076.00	8 025.00	12 631.08	841.543 6
Carbon (dry) (%)	79.32	44.03	70.43	5.026 348
Hydrogen (dry) (%)	5.36	3.39	4.78	0.310 245
Nitrogen (dry) (%)	3.03	0.35	1.40	0.290 988
Oxygen (dry) (%)	12.57	2.16	7.53	1.660 288
HGI	72.00	30.00	58.80	5.710 457

## Evaluation of the effect of coal chemical properties on the Hardgrove Grindability Index

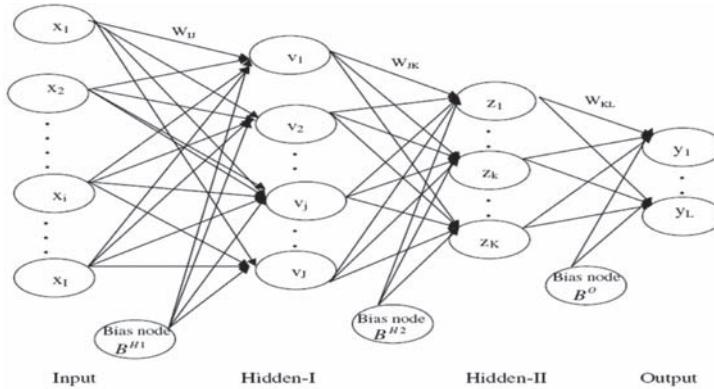


Figure 1—MLP architecture with two hidden layers (Patel *et al.*, 2007)

This algorithm is more powerful than the commonly used gradient descent methods, because the Levenberg-Marquardt approximation makes training more accurate and faster near minima on the error surface (Lines and Treitel, 1984).

The method is as follows:

$$\Delta W = (J^T J + \mu I)^{-1} J^T e \quad [1]$$

In Equation [1] the adjusted weight matrix  $\Delta W$  is calculated using a Jacobian matrix  $J$ , a transposed Jacobian matrix  $J^T$ , a constant multiplier  $\mu$ , a unity matrix  $I$  and an error vector  $e$ . The Jacobian matrix contains the weights derivatives of the errors:

$$J = \begin{vmatrix} \frac{\partial E}{\partial w_{ij}} & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \frac{\partial E}{\partial w_m} \end{vmatrix} \quad [2]$$

If the scalar  $\mu$  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 (Haykin, 1999). For more detailed information the reader is referred to Lines and Treitel, 1984.

After each successful step (lower errors) the constant  $\mu$  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  $\mu$  is increased subsequently. The weights of the adjusted weight matrix (Equation [2]) are used in the forward pass. The mathematics of both the forward and backward pass is briefly explained in the following paragraphs.

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 HGI value of sample) are calculated by:

$$net_{pj} = \sum_{n=1}^{last} w_{jn} O_{pn} \quad [3]$$

where the number of neurons in the previous layer ( $L-1$ ) is defined by  $n=1$  to the 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:

$$O_{pj} = f_{pj}(net_{pj}) = \frac{1}{1 + e^{-(net_{pj} + \theta_j)}} \quad [4]$$

where  $\theta_j$  is the bias of neuron  $j$ .

In general the output vector, containing all  $O_{pj}$  of the neurons of the output layer, is not the same as the true output vector (i.e. the measured HGI 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:

$$E_p = \frac{1}{2} \sum (t_{pj} - O_{pj})^2 \quad [5]$$

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 (Haykin, 1999) as:

$$E_t = \sum E_p \quad [6]$$

To reduce the training error, the connection weights are changed during a completion of a forward and backward pass by adjustments ( $\Delta w$ ) of all the connections weights  $w$ . Equations [2] and [3] 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 (Haykin, 1999).

### Training and testing of the model

As mentioned, the input layer has six neurons  $X_i$ ,  $i=1, 2, \dots, 6$ . The number of neurons in the hidden layer is supposed  $Y$ , the output of which is categorized as  $P_j$ ,  $j_{pj}=1, 2, \dots, Y$ . The output layer has one neuron which denotes the amount of gold extraction. It is assumed that the connection weight matrix between input and hidden layers is  $W_{ij}$ , and the connection weight matrix between hidden and output layers is  $W_{HK}$ ,  $K$  denotes the learning sample numbers. A schematic presentation of the whole process is shown in Figure 2.

Nonlinear (LOGSIG, TANSIG) and linear (PURELIN) functions can be used as transfer functions (Figures 3 and 4). The logarithmic sigmoid function (LOGSIG) is defined as (Demuth and Beale, 1994):

## Evaluation of the effect of coal chemical properties on the Hardgrove Grindability Index

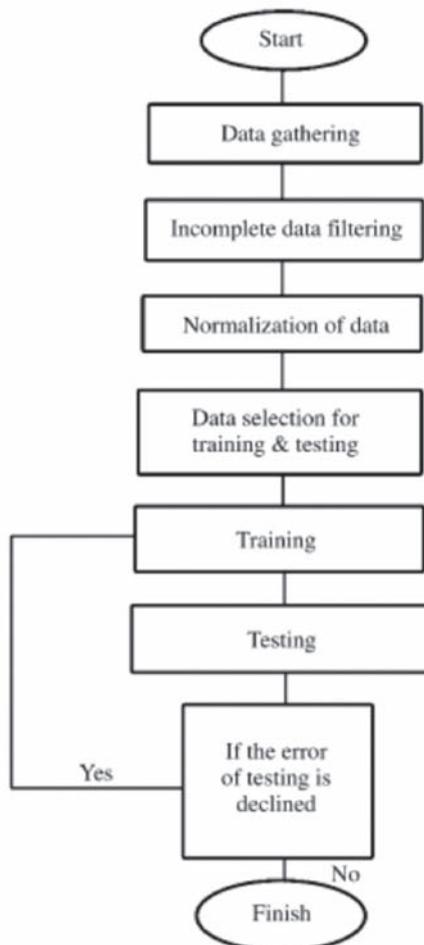


Figure 2—ANN process flow chart

$$f = \frac{1}{(1 + e^{-e_x})} \quad [7]$$

whereas, the tangent sigmoid function (TANSIG) is defined as follows (Demuth and Beale, 1994):

$$f = \frac{e^{e_x} - e^{-e_x}}{e^{e_x} + e^{-e_x}} \quad [8]$$

where  $e_x$  is the weighted sum of the inputs for a processing unit.

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 were examined (Table II).

Multilayer network architecture with two hidden layers between the input and output units is applied. During the design and development of the neural network for this study, it was determined that a four-layer network with 10 neurons in the hidden layers (two layers) would be the most appropriate. The ANN architecture for predicting the HGI is shown in Figure 5.

Table II Results of a comparison between some of the models			
No.	Transfer function	Model	SSE
1	LOGSIG-LOGSIG	10-4-1	1.5
2	LOGSIG-LOGSIG-LOGSIG	10-8-7-1	1.2
3	TANSIG-LOGSIG-LOGSIG	10-5-5-1	0.9
4	TANSIG-LOGSIG-LOGSIG	10-10-4-1	0.4
5	TANSIG-LOGSIG-PURELINE	10-14-5-1	0.25
6	TANSIG-TANSIG-PURELINE	10-5-5-1	0.016

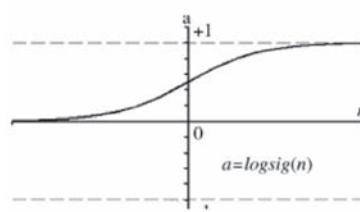


Figure 3—Sigmoid transfer functions

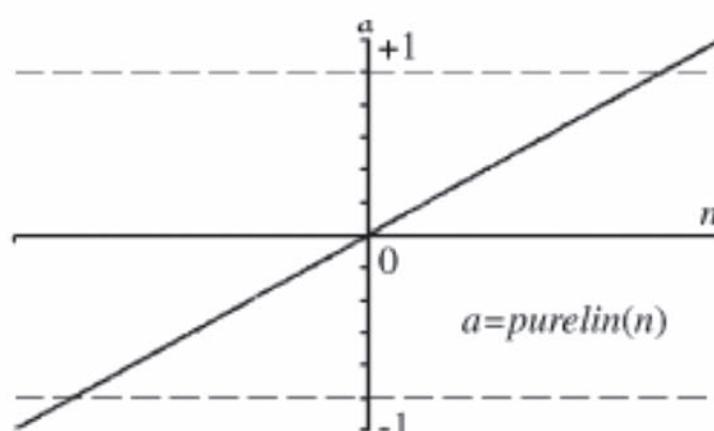
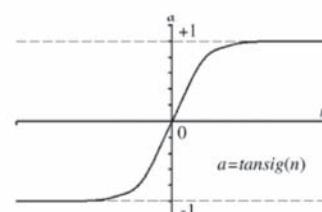


Figure 4—Linear transfer function

## Evaluation of the effect of coal chemical properties on the Hardgrove Grindability Index

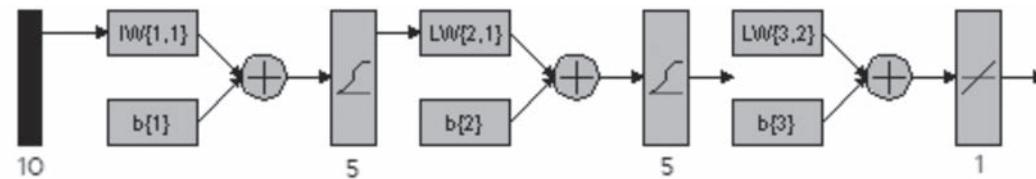


Figure 5—ANN architecture for predict the HGI

The learning rate of the network was adjusted so that 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 to avoid overtraining, which would cause the 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 Figure 6. As shown, the optimum training was achieved at about 200 epochs.

For the evaluation of a model, the predicted and measured values of HGI can be compared. For this purpose, MAE ( $E_a$ ) and mean relative error ( $E_r$ ) can be used.  $E_a$  and  $E_r$  are computed as follows (Demuth and Beale, 1994):

$$E_a = |T_i - O_i| \quad [9]$$

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

where  $T_i$  and  $O_i$  represent the measured and predicted outputs.

For the optimum model,  $E_a$  and  $E_r$  were equal to 0.503 and 0.0125 respectively. A correlation between the measured and predicted HGI for training and testing data is shown in Figures 7 and 8 respectively. It can be seen that the coefficient of correlation in both of the processes is very good.

### Sensitivity analysis

To analyse the strength of the relationship between the backbreak and the input parameters, the cosine amplitude method (CAM) was utilized. The CAM was used to obtain the express similarity relations between the related parameters. To apply this method, all of the data pairs were expressed in common  $X$ -space. The data pairs used to construct a data array  $X$  were defined as (Demuth and Beale, 1994):

$$X = \{X_1, X_2, X_3, \dots, X_m\} \quad [11]$$

Each of the elements,  $X_i$ , in the data array,  $X$ , is a vector of lengths of  $m$ , that is:

$$X_i = \{x_{i1}, x_{i2}, x_{i3}, \dots, x_{im}\} \quad [12]$$

Thus, each of the data pairs can be thought of as a point in  $m$ -dimensional space, where each point requires  $m$ -coordinates for a full description. Each element of a relation,  $r_{ij}$ , results in a pairwise comparison of two data pairs. The strength of the relation between the data pairs,  $x_i$  and  $x_j$ , is given by the membership value expressing the strength:

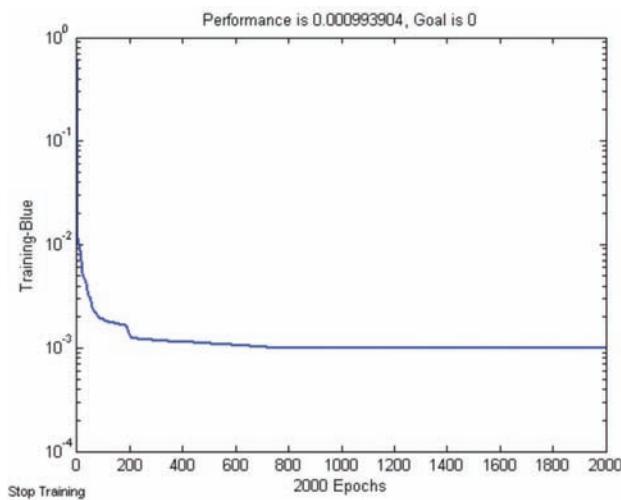


Figure 6—Network performance during the training process

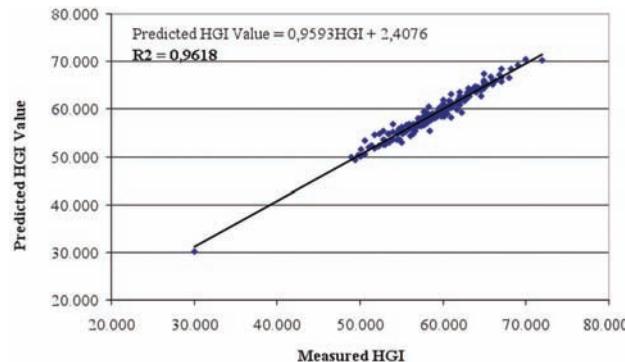


Figure 7—Correlation between measured and predicted HGI for training data

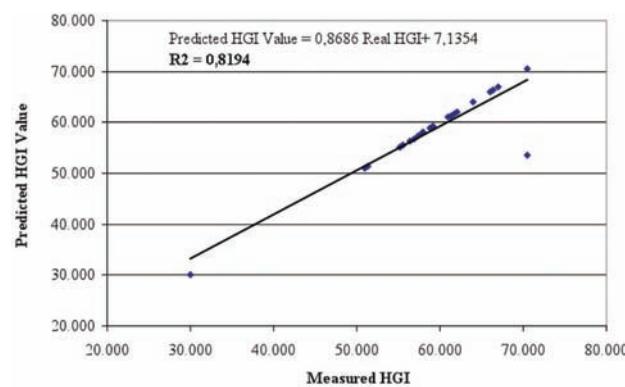


Figure 8 – Correlation between measured and predicted HGI for testing data

## Evaluation of the effect of coal chemical properties on the Hardgrove Grindability Index

$$r_{ij} = \sum_{k=1}^m x_{ik} x_{jk} / \sqrt{\sum_{k=1}^m x_{ik}^2 \sum_{k=1}^m x_{jk}^2} \quad [13]$$

The strengths of the relations ( $r_{ij}$  values) between HGI and input parameters (coal chemical properties) are shown in Figure 9. As can be seen, the effective parameters on HGI include the volatile matter (dry), Btu/lb (dry), carbon (dry), hydrogen (dry), fixed carbon (dry), nitrogen (dry), oxygen (dry), moisture, ash (dry), and total sulphur (dry). It is possible to consider and examine the effective parameters in the coal HGI, and modification was also applied by changing the further effective parameter.

### Discussion

In this investigation the effect of coal chemical properties on the HGI were studied. Results from the neural network showed that volatile matter (dry), Btu (dry), and carbon (dry) were the parameters with the most effect on HGI, respectively. The least effective input parameters on HGI were total sulphur (dry) and ash (dry) respectively. Figures 7 and 8 shows that the measured and predicted HGI values are similar. The results of the ANN show that the correlation coefficients ( $R^2$ ) achieved for training and test data were 0.9618 and 0.8194 respectively.

### Conclusion

In this research, an artificial neural network approach was used to evaluate the effects of chemical properties of coal on HGI. Input parameters were moisture, volatile matter (dry), fixed carbon (dry), ash (dry), total sulphur (organic and pyretic) (dry), Btu/lb (dry), carbon (dry), hydrogen (dry), nitrogen (dry), and oxygen (dry). According to the results, the optimum ANN architecture has been found to be five and five neurons in the first and second hidden layer, respectively, and one neuron in the output layer. In the ANN method, the correlation coefficients ( $R^2$ ) for the training and test data were 0.9618 and 0.8194, respectively.

Sensitivity analysis of the network shows that the most effective parameters influencing the HGI were volatile matter (dry), Btu/lb (dry), carbon (dry), hydrogen (dry), fixed carbon (dry), nitrogen (dry) and oxygen (dry), respectively, and those with the least effect were moisture, ash (dry), and total sulphur (dry), respectively (Figure 9).

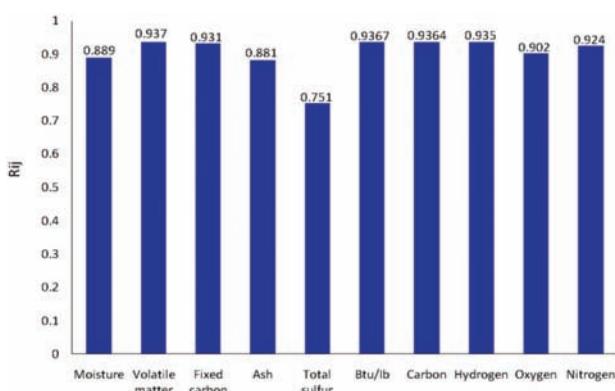


Figure 9—Strengths of relation ( $r_{ij}$ ) between HGI and each input parameter

As regards network training performance, the error of the training network minimized when the number of epochs was 200, and after this point the best performance was achieved for the network. The values of  $E_a$  and  $E_r$  from the ANN were 0.503 and 0.0125, respectively.

### References

GÜLHAN ÖZBAYOĞLU, A., MURAT ÖZBAYOĞLU, M., and EVREN ÖZBAYOĞLU. 2008. Estimation of Hardgrove Grindability Index of Turkish coals by neural networks. *International Journal of Mineral Processing*, vol. 85, no. 4. pp. 93–100.

LITTLE, J., CHOI, N., and PRISBREY, K. 1992. Influence of preheating on grindability of coal. *International Journal of Mineral Processing*, vol. 36, no. 1–2. pp. 107–112.

URAL, S. and AKYILDIZ, M. 2004. Studies of relationship between mineral matter and grinding properties for low-rank coal. *International Journal of Coal Geology*, vol. 60, no. 1. pp. 81–84.

JORJANI, E., HOWER, J.C., CHEHREH CHELGANI, S., SHIRAZI, MOHSEN A., and MESROGHLI, SH. 2008. Studies of relationship between petrography and elemental analysis with grindability for Kentucky coals. *Fuel*, vol. 87, no. 6. pp. 707–713.

VUTHALURU, H.B., BROOKE, R.J., ZHANG, D.K., and YAN, H.M. 2003. Effects of moisture and coal blending on Hardgrove Grindability Index of Western Australian coal. *Fuel Processing Technology*, vol. 81, no. 1. pp. 67–76.

TIRYAKI, B. 2005. Technical note: Practical assessment of the grindability of coal using its hardness characteristics. *Rock Mechanics and Rock Engineering*, vol. 38, no. 2. pp. 145–151.

PARASHER, C.L. 1987. *Crushing and Grinding Process Handbook*. Wiley, Chichester, UK. pp. 216–227.

YAO, H.M., VUTHALURU, H.B., TADE, M.O., and DJUKANOVIC, D. 2005. Artificial neural network-based prediction of hydrogen content of coal in power station boilers. *Fuel*, vol. 84, no. 12–13. pp. 1535–1542.

HAYKIN, S. 1999. *Neural Networks. A Comprehensive Foundation*. 2nd edn. Prentice Hall.

UNGAR, L.H., HARTMAN, E.J., KEELER, J.D., and MARTIN, G.D. 1996. Process modeling and control using neural networks. *American Institute of Chemical Engineering, Symposium Series*, vol. 92. pp. 57–66.

PLIPPEN, R. 1987. An introduction to computing with neural nets. *IEEE ASSP Magazine*, vol. 4. pp. 4–22.

KHOSHJAVAN, S., REZAI, B., and HEIDARY, M. 2011. Evaluation of effect of coal chemical properties on coal swelling index using artificial neural networks. *Expert Systems with Applications*, vol. 38, no. 10. pp. 12906–12912.

ZEIDENBERG, M. 1990. *Neural Network Models in Artificial Intelligence*. E. Horwood, New York. p. 16.

PATEL, S.U., KUMAR, B.J., BADHE, Y.P., SHARMA, B.K., SAHA, S., BISWAS, S., CHAUDHURY, A., TAMBE, S.S., and KULKARNI, B.D. 2007. Estimation of gross calorific value of coals using artificial neural networks. *Fuel*, vol. 86, no. 3. pp. 334–344.

LINES, L.R. and TREITEL, S. 1984. A review of least-squares inversion and its application to geophysical problems. *Geophysical Prospecting*, vol. 32.

DEMUTH, H. and BEALE, M. 1994. *Neural Network Toolbox User's Guide*. The MathWorks Inc., Natick, MA. ♦