Scientific Research and Essays Vol. 5(16), pp. 2242-2249, 18 August, 2010
Available online at http://www.academicjournals.org/SRE
ISSN 1992-2248 ©2010 Academic Journals

Full Length Research paper

Different types of learning algorithms of artificial neural network (ANN) models for prediction of gross calorific value (GCV) of coals

Işik Yilmaz[1*], Nazan Yalcin Erik[1] and Oğuz Kaynar[2]

[1]Department of Geological Engineering, Faculty of Engineering, Cumhuriyet University, 58140 Sivas, Turkey.
[2]Department of Management Information System, Cumhuriyet University, 58140 Sivas, Turkey.

Accepted 15 July, 2010

Correlations are very significant from earliest days, in some cases, it is essential as it is difficult to measure the amount directly, and in other cases, it is desirable to ascertain the results with other tests through correlations. Soft computing techniques are now being used as alternative statistical tools, and new techniques such as; artificial neural networks, fuzzy inference systems, genetic algorithms, etc. and their hybrid forms have been employed for developing of the predictive models to estimate the needed parameters, in the recent years. Determination of gross calorific value (*GCV*) of coals is very important to characterize coal and organic shales; it is difficult, expensive, time consuming and is a destructive analysis. In this paper, use of different learning algorithms of artificial neural networks such as MLP, RBF (exact), RBF (k-means) and RBF (SOM) for prediction of *GCV* was described. As a result of this paper, all models exhibited high performance for predicting *GCV*. Although the four different algorithms of ANN have almost the same prediction capability, accuracy of MLP has relatively higher than other models. The use of soft computing techniques will provide new approaches and methodologies in prediction of some parameters in the investigations about the fuels.

Key words: ANN, MLP, RBF, soft computing, coal, gross calorific value.

INTRODUCTION

The goal of coal mining is to economically remove coal from the ground. Coal is valued for its energy content, and since the 1880s, it is widely used to generate electricity. Steel and cement industries use coal as a fuel for the extraction of iron from iron ore and also for cement production. Coal mining is one of the most important mining activities in the world and they are found largely in terrestrial areas in different parts of the world (Marschalko, 2008, 2009). Coal is an important and prevalent fuel in the world, which supplies ~40% of the planet's energy needs. Further development of coal production, combustion and emission control technologies can ensure that coal continues to contribute to energy security, economic growth, and environmental protection. Turkey has also large coal reserves –about 9 GT (Turkish Lignite Authority, 2008). The low-rank coals of Turkey represent the country's major energy source with their relatively large geological reserves. The coal-bearing terrestrial Tertiary deposits of Turkey overlie an area of approximately 110000 km^2 with the thicknesses of the coal seams varying from 0.05 - 87 m. The country's total coal reserves are 8 - 13 billion tons of low-quality coal, and 13 - 40 million tons of bituminous coal. Of the total annual lignite production of 90 million tons, 80% is consumed by thermal power plants. The current total amount of coal consumption of the country, including imported coals as well, is about 60 MT each year. Power plants and iron-steel plants consume most of this coal and cause the emission of high amounts of CO_2 each year.

Soft computing techniques such as Fuzzy logic, artificial neural networks, genetic algorithms and neuro-Fuzzy systems, are generally used in design of higher technology products and are now being used in different

*Corresponding author. E-mail: iyilmaz@cumhuriyet.edu.tr. Tel: +90 346 219 1010 (1305 ext.). Fax: +90 346 219 1171.

Table 1. Basic statistics of the results obtained from analyses.

	GCV	C	H	N	O	S
Minimum	2.12	29.49	1.37	0.21	4.23	0.13
Maximum	27.65	78.52	6.56	2.54	67.63	9.12
Average	14.47	59.69	4.23	1.17	31.36	3.54
Std. Dev.	6.461	13.660	1.226	0.582	16.853	2.208

Unit of GCV is MJ/kg, others are %.

Table 2. Descriptive statistics for GCV as an independent value.

N	Valid :74
	Missing: 0
Mean	14.4711
Std. error of mean	0.7510
Median	14.4750
Mode	9.89
Std. deviation	6.4607
Variance	41.7410
Skewness	0.028
Std. error of skewness	0.279
Kurtosis	0.931
Std. error of Kurtosis	0.552
Range	25.54
Minimum	2.12
Maximum	27.66
Sum	1070.86

branches of sciences and technologies, and their popularities has gradually been increased.

Earth sciences aim to describe very complex processes, and are needed in new technologies for data analyses. The number of researches in evolutionary algorithms and genetic programming, neural science and neural net systems, Fuzzy set theory and Fuzzy systems, fractal, chaos theory and chaotic systems, aiming to provide solution to the problems in earth sciences (estimation of parameters; susceptibility, risk, vulnerability and hazard mapping; interpretation of geophysical measurement results; many kinds of mining applications; etc.) has been increased in the last five to ten years, particularly.

Correlations have been a significant part of scientific researches from earliest days. The correlations are generally semi-empirical based on some mechanics or purely empirical that is based on statistical analysis (Yilmaz, 2006). However, determination of gross calorific value (GCV) of a coal is time consuming, expensive and involves destructive tests. If reliable predictive models could be obtained between GCV with quick, cheap and nondestructive test results, it would be very valuable for the characterization of coals.

Ultimate analyses characterize the chemical composition of coals and measure various element contents such as; carbon (C), hydrogen (H), nitrogen (N), sulphur (S), and oxygen (O). This study aims to predict the GCV of coals using different learning algorithms of ANN. A total of 74 coal samples were collected from various locations in Turkey and ultimate analyses were performed. These parameters were correlated with GCV first and statistically significant ones were selected. In order to establish predictive models, different learning algorithms of artificial neural networks (ANNs) such as MLP, RBF (exact), RBF (k-means) and RBF (SOM) were used and prediction performances were then analyzed.

TESTED COALS AND DATA SET

For ultimate analyses, selected coal samples were purified according to American Society for Testing and Materials (ASTM) guidelines (ASTM 2004 D5373). These analyses were performed in the Turkish General Directorate of Mineral Research and Exploration laboratory (MTA-MAT Laboratory, Ankara, Turkey) using standard analytical procedures. Sulfur (wt.%, *daf*), carbon (wt.%, *daf*), hydrogen (wt.%, *daf*) and nitrogen (wt.%, *daf*) contents were determined using a LECO analyzer in the same laboratory. The results obtained and their basic test statistics are tabulated in Table 1. The gross calorific value (GCV) of the coals ranged between 2.12 and 27.65 MJ/kg with an average value of 14.47 MJ/kg.

More attention should be paid particularly, in selecting the data set with a normal distribution. In order to characterize the variation of GCV used as an independent value, descriptive statistics such as; minimum, maximum, mean, mode, median, variance, standard deviation, skewness and kurtosis etc. were calculated using the SPSS Version 10.0 (1999) package. Table 2 presents the independent values which show almost normal distribution. However, it is close to the normal distribution, data are skewed right and showed kurtosis (Figure 1). It can be seen that the respective skewness and kurtosis values of 0.028 and 0.931 were very low. In conclusion, it is suitable that the regression analyses will work well in this case.

In order to establish the predictive models among the parameters obtained in this study, simple regression analyses were performed in the first stage. The relations between GCV with other parameters, were analyzed by employing linear, power, logarithmic and exponential functions. Statistically significant and strong correlations were then selected (Table 3). All obtained relationships were found to be statistically significant according to the Student's *t*-test with 95% confidence, except nitrogen (N). Figure 2 shows the plot of the GCV versus carbon (C), hydrogen (H), nitrogen (N), sulphur (S) and oxygen (O).

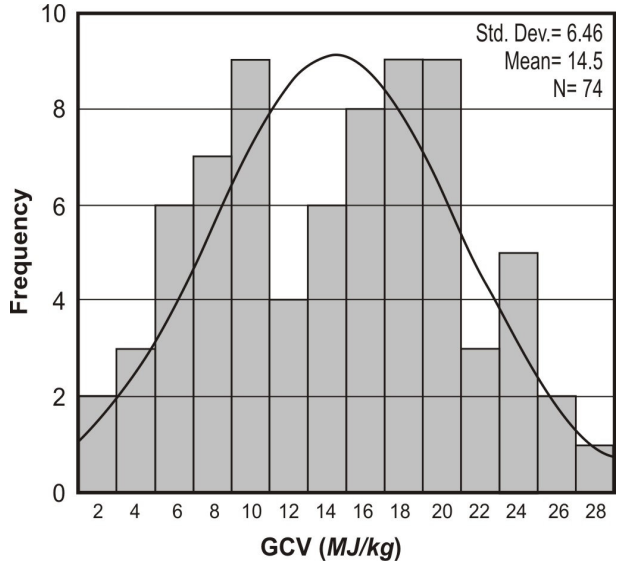

Figure 1. Frequency distribution of GCV values of samples used in analyses.

Table 3. Correlation coefficients (R^2) obtained from the simple regressions between GCV with other parameter.

Model	Set – II (Independents)				
	C	H	N	O	S
Linear	0.71	0.62	0.29	0.75	0.66
Logarithmic	0.67	0.60	0.29	0.72	0.64
Exponential	0.70	0.62	0.31	0.74	0.58
Power	0.69	0.68	0.34	0.62	0.72

Gray filled cells show the highest correlation coefficients (R^2), gray filled cells with borders show strong correlations which were included in the models.

AN OVERVIEW OF ARTIFICIAL NEURAL NETWORK (ANN) MODELS

When the materials are natural, there is always an uncertainty sourced from the nature of the materials. This could be the main reason why soft computing methodologies such as artificial neural networks, Fuzzy systems, genetic algorithms have been developed in recent years. These techniques attract more and more attention in several research fields because they tolerate a wide range of uncertainty (Jin and Jiang, 1999).

Artificial neural networks are data processing systems devised via imitating brain activity and have performance characteristics like biological neural networks. ANN has a lot of important capabilities such as learning from data, generalization and working with unlimited number of variable (Kaynar et al., 2011). Neural networks may be used as a direct substitute for auto correlation, multivariable regression, linear regression, trigonometric and other statistical analysis and techniques (Singh et al., 2003). Neural networks, with their remarkable ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other

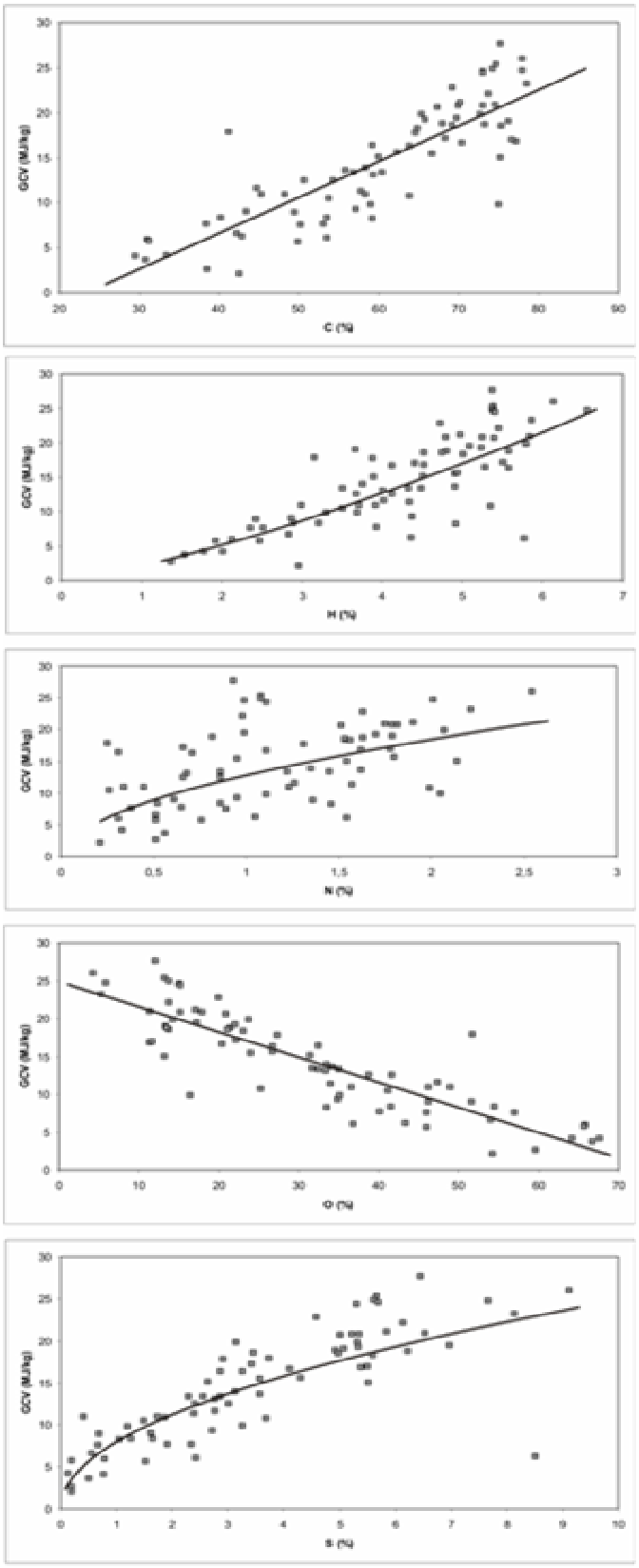

Figure 2. GCV versus carbon (C), hydrogen (H), nitrogen (N), oxygen (O), sulphur (S) graphs.

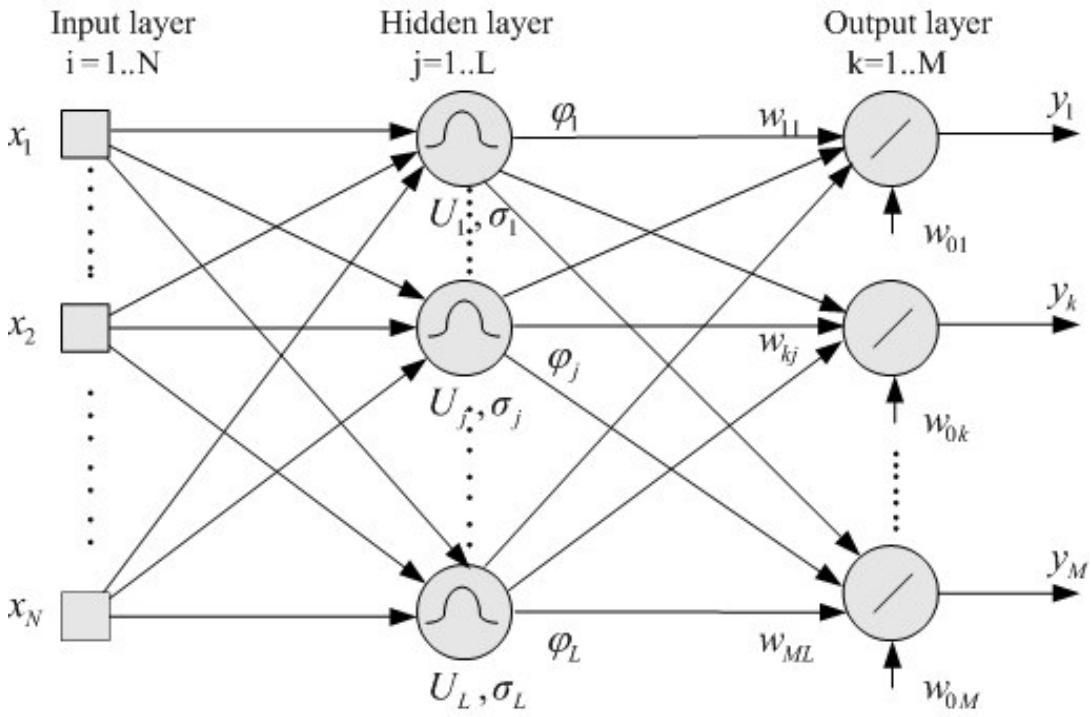

Figure 3. Architecture of radial basis function network (RBF).

computer techniques (Yilmaz, 2009a, b). Rumelhart and McClelland (1986) reported that the main characteristics of ANN include large-scale parallel distributed processing, continuous nonlinear dynamics, collective computation, high fault-tolerance, self-organization, self-learning, and real-time treatment. A trained neural network can be thought of as an "expert" in the category of information given to be analyzed. This expert can then be used to provide projections that given new situations of interest and answer "what if" questions (Yilmaz, 2010).

The most commonly used algorithms are multilayer feed forward artificial neural network (Multiple Layer Perceptron-MLP) and radial basis function networks (RBFN). The radial basis function network (RBFN) is traditionally used for strict interpolation problem in multi-dimensional space, and has similar capabilities with MLP neural network which solves any function approximation problem (Park and Sandberg, 1993). RBFs were first used in designing neural network by Broomhead and Lowe (1988), who showed how a nonlinear relationship could be modelled by RBF neural network, and interpolation problems could be implemented (Broomhead and Lowe, 1988). The main two advantages of RBFN are:

1. Training of networks in a short time than MLP (Moody and Darken, 1989),
2. Approximation of the best solution without dealing with local minimums (Park and Sandberg, 1991).

Moreover, RBFN are local networks compared to the feed-forward networks which perform global mapping. Otherwise, RBFN uses a single set of processing units, and each of these units is most receptive to a local region of the input space (Xu et al., 2003). That is why, RBFN are used as an alternative neural network model in applications of function approximation, time series forecasting as well as classifying task in recent years (Chen et al., 1991; Bianchini et al., 1995; Sheta and Jong, 2001; Foody, 2004; Rivas et al., 2004; Harpham and Dawson, 2006; Sarimveis et al., 2006; Zhang et al., 2007; Yu et al., 2008).

The structure of RBFN is composed of three layers (Figure 3), and the main distinction between MLP and RBFN is the number of the hidden layer. RBFN has only one hidden layer which contains nodes called RBF units, and radially symmetric basis function is used as activation functions of hidden nodes.

The input layer serves as an input distributor to the hidden layer. Differently from MLP, the values in input layer are forwarded to hidden layer directly without being multiplied by weight values. The hidden layer unit measures the distance between an input vector and the centre of its radial function, and produces an output value depending on the distance. The centre of radial basis function is called reference vector. The closer the input vector is to the reference vector, the more the value is produced at the output of hidden node. However, a lot of radial basis functions are suggested for using in hidden layer (Gausian, Multi-Quadric, Generalized Multi-Quadric, Thin Plate Spline), Gaussian function is the most widely used in applications. Chen et al. (1991) indicate that the choice of radial basis function used in network does not significantly affect performance of network. The activation function of the individual hidden nodes defined by the Gaussian function is expressed as follows:

$$\varphi_j = e^{\left[-\frac{\|X-C_j\|^2}{\sigma_j^2}\right]} \qquad j = 1,2....,L \qquad (3)$$

where φ_j denotes the output of the j th node in hidden layer, $\|.\|$ is Euclidian distance function which is generally used in applications, X is the input vector, C_j is center of the j th

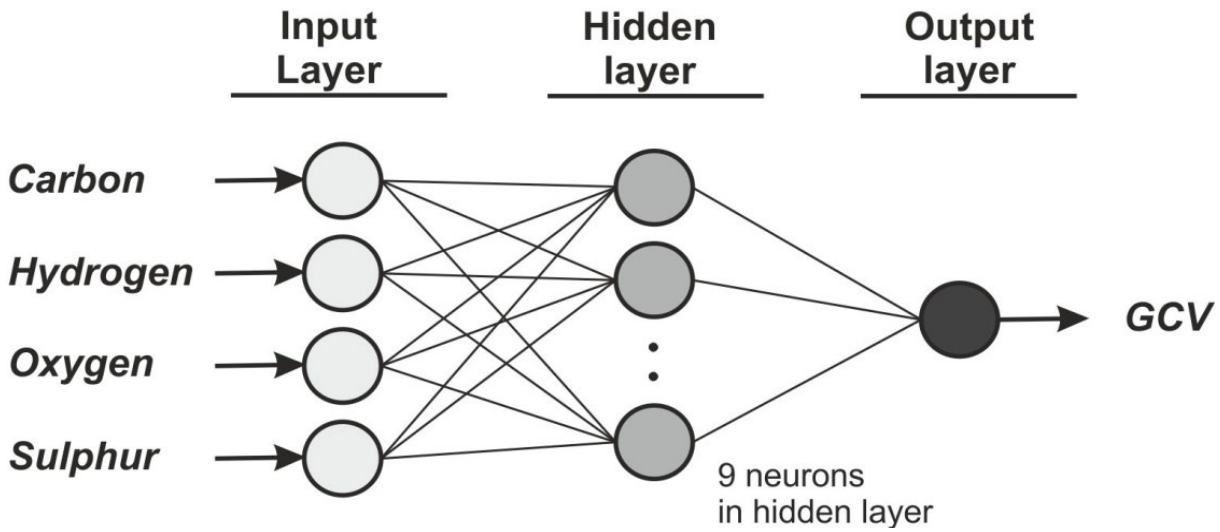

Figure 4. MLP model used in this study.

Gaussian function, σ_j is radius which shows the width of the Gaussian function of the j th node and L denotes the number of hidden layer nodes.

In the next step, the neurons of the output layer perform a weighted sum using the hidden layer outputs and the weights which connect hidden layer to output layer. Output of network can be presented as a linear combination of the basis functions:

$$y_k = \sum_{j=1}^{L} \varphi_j w_{kj} + w_{k0} \quad (4)$$

Where: w_{kj} is the weight that connects hidden neuron j and output neuron k , w_{k0} is bias for the output neuron.

ARTIFICIAL NEURAL NETWORK MODELS FOR PREDICTION OF GCV

All data were first normalized and divided into three data sets such as; training (1/2 of all data), test (1/4 of all data), and verification (1/4 of all data). In this study MatLab 7.1 (2005) software was used in neural network analyses having a three-layer feed-forward network, models were constructed by MLP and RBF architectures.

MLP model for prediction of GCV

In this study, gross calorific value of coals was first predicted indirectly by using the MLP algorithm. They consist of an input layer (4 neurons), one hidden layer (9 neurons) and one output layer (Figure 4). In the analyses network parameters of learning parameter, momentum parameters and networks training function, which is an activation (transfer) function for all layers, were respectively adjusted to 0.01, 0.9, trainLm and tansig. As in many other network training methods, models and parameters were used in order to reach the minimum RMS values, and network goal was reached at the end of 537 iterations.

In fact, the coefficient of determination between the measured and predicted values is a good indicator to check the prediction performance of the model. Figure 5a shows the relationships between measured and predicted values obtained from the models for GCV, with good coefficient of determinations. In this study, variance account for VAF (Equation 5) and root mean square error (RMSE) (Equation 6), indices were also calculated to control the performance of the prediction capacity of predictive models developed in the study as employed by Alvarez and Babuska (1999), Finol et al. (2001) and Yilmaz and Yüksek (2008, 2009):

$$VAF = \left[1 - \frac{\mathrm{var}\left(y - y'\right)}{\mathrm{var}\left(y\right)}\right] \times 100 \quad (5)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y - y'\right)^2} \quad (6)$$

Where; y and y' are the measured and predicted values, respectively. If the VAF is 100 and RMSE is 0, then the model will be excellent. The obtained values of VAF and RMSE given in Table 4 indicated a high prediction performance.

RBF models for prediction of GCV

Training of RBF networks contains process of determination of centre vector (C_j), radius value (σ_j) and linear weight values (w_{kj}). Two stage hybrid learning algorithm is used to train RBF networks in general. In the first stage of hybrid learning algorithm, centre and width of RBFs in hidden layer are determined by using unsupervised clustering algorithms or randomly selected from given input data set. Output weight is calculated in the second stage. A lot of methods are proposed in literature to determine centre and width of reference vector, and some of them are listed.

Number of hidden neurons is set to the number of training examples and all input vectors are also used as centres of RBFs. In other words, for each point in input space, one radial basis function is determined. This case is named as “Exact RBF”. There

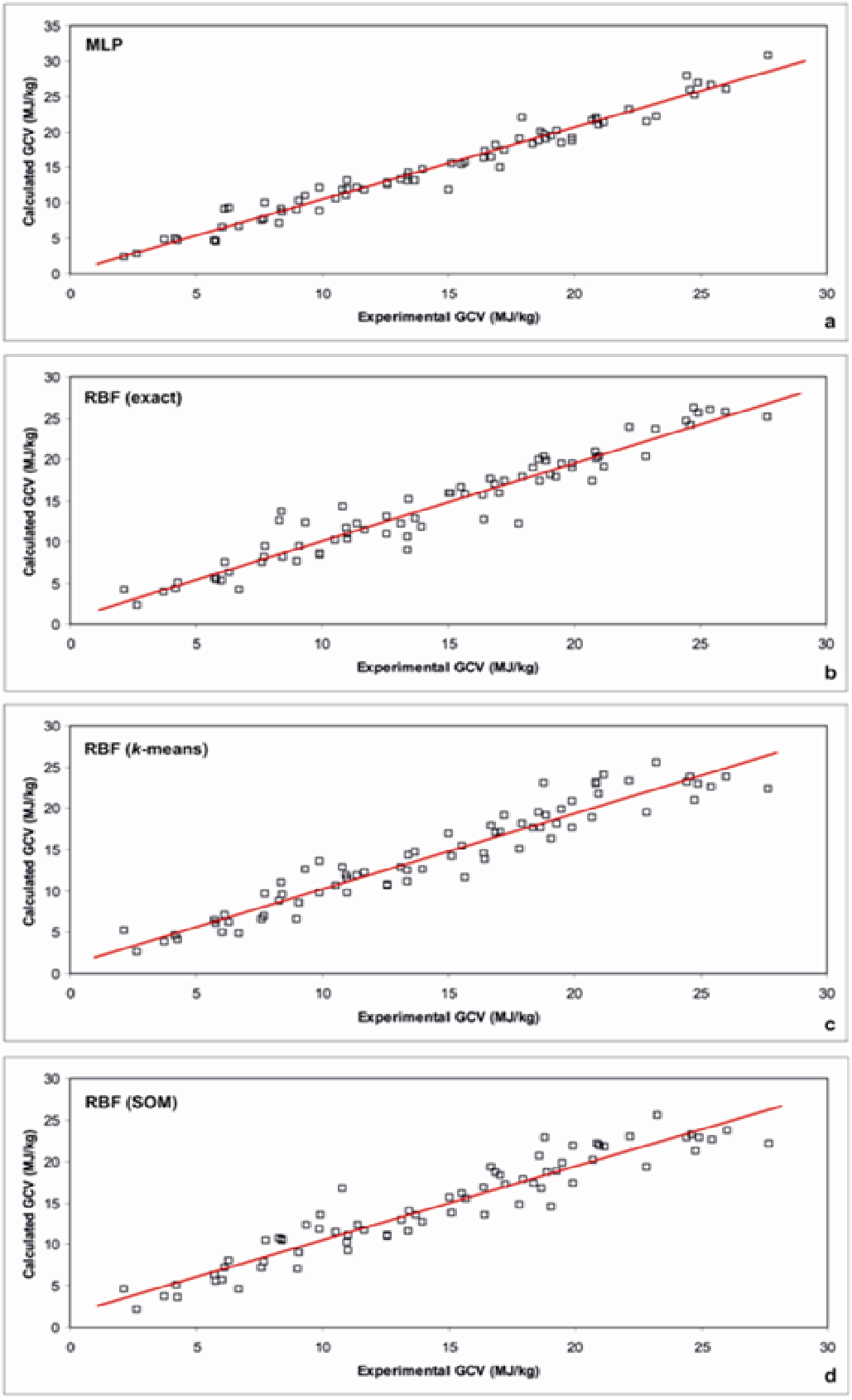

Figure 5. Cross-correlation of predicted and observed values of *GCV* for: (a) MLP, (b) Exact RBF, (c) RBF (k-means), (d) RBF (SOM) models.

Table 4. Performance indices (RMSE, VAF and R^2) for models used.

Model	RMSE	VAF (%)	(R^2)
ANN- MLP	1.341	96.584	0.966
ANN- RBF (exact)	3.063	93.234	0.926
ANN- RBF (k-means)	3.534	91.256	0.914
ANN- RBF (SOM based)	4.024	89.993	0.903

RMSE = root mean square error, VAF= value account for.

are two disadvantages of Exact RBF such as; size problem and overtraining problem. Size problem causes calculation complexity when data set is too large. Network is over trained with these noisy data, so, performance of the system for test data will not as lucrative as performance of training data. To reduce the calculating complexity and to deal with overtraining problem, the number of neurons in hidden layer is reduced as smaller than the number of sample in input data set. And, central vectors are chosen from input vectors, randomly.

Pruning or growing methods, which start iteratively with a number of pre-specified hidden neuron, continues by adding/removing hidden neurons to/from the RBFN. The network structure which has minimum testing and training error is selected as a final model of RBFN. In this iterative process, parameters of hidden nodes are randomly selected from input vectors or determined by using clustering methods. In order to determine central vectors with clustering methods, input vectors are devoted to certain number of clusters by using clustering algorithms such as; *k*-means, Self Organization Map (SOM), etc., cluster centres are then used as RBF centres.

In the analyses, three different algorithms of RBF such as; Exact RBF, RBF trained with *k*-means and RBF trained with SOM were used in the prediction of GCV. However, three models consist of 4 neurons in input layer and one output layer, the neuron numbers of in the hidden layer of Exact RBF, RBF trained with *k*-means, RBF trained with SOM were 42, 27 and 30, respectively.

Cross-correlation between predicted and observed values (Figures. 5b, c, d) and RMSE and VAF values indicated that the three models of RBF constructed is highly acceptable for predicting GCV. RMSE, VAF and R^2 values are also tabulated in Table 4.

RESULTS AND CONCLUSION

In this paper, utilization of some neural computing models such as; artificial neural network (ANN) with different algorithms (MLP, Exact RBF, RBF trained with *k*-means and RBF trained with SOM) for prediction of gross calorific value of coals was described and compared. It appears that there is a possibility of estimating GCV of coals from ultimate analyses results by using the soft computing models.

The results of the present paper showed that ANN models having four inputs and one output can be applied successfully for prediction of GCV, and models exhibited reliable predictions. However, all four different algorithms of ANN has almost the same prediction capability, accuracy of MLP was relatively higher than RBF models. The comparison of VAF, RMSE indices and coefficient of correlations (R^2) for predicting GCV, revealed that

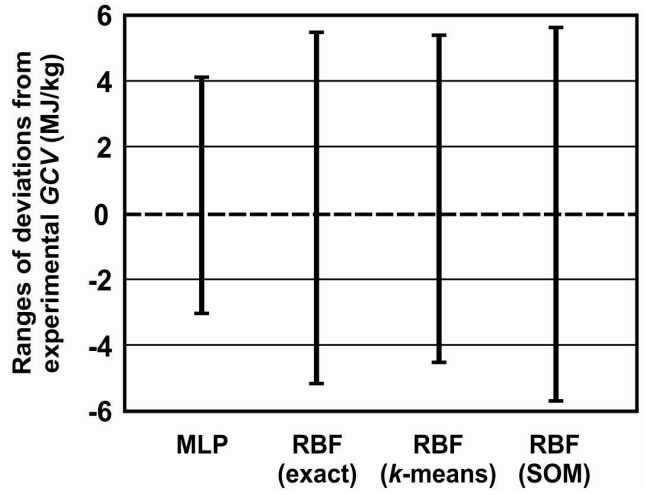

Figure 6. Intervals of variations of predicted from observed values.

prediction performances of MLP model is higher than those of other algorithms of artificial neural networks; Exact RBF, RBF trained with *k*-means and RBF trained with SOM. In order to show the deviations from the observed values of GCV, the distances of the predicted values from the models constructed from the observed values were also calculated and it was found that the deviation intervals (-3.16 - +4.13) of the predicted values from MLP are smaller than the deviation intervals of ANN models (in Exact RBF -5.28 - +5.58, in RBF trained with *k*-means -4.28 - +5.27, in RBF trained with SOM -5.92 - +5.44) (Figure 6).

As it is known, the potential benefits of soft computing models extend beyond the high computation rates. Higher performances of the soft computing models were sourced from greater degree of robustness and fault tolerance than traditional statistical models, because there are many more processing neurons, each with primarily local connections.

The performance comparison also showed that the soft computing techniques are good tools for minimizing the uncertainties, and their utilization may provide new approaches and methodologies, and minimize the potential inconsistency of correlations. The results of this paper will provide dissemination of important results of the use of soft computing technologies in fuel sciences, and serve as an example for fuel geologists and engineers engaged in this area of interest

REFERENCES

Alvarez GM, Babuska R (1999). Fuzzy model for the prediction of unconfined compressive strength of rock samples. Int. J. Rock Mech. Min. Sci. 36: 339-349.

American Society for Testing and Materials (ASTM) D5373 (2004). Standard test methods for instrumental determination of carbon, hydrogen and nitrogen in laboratory samples of coal and coke. In:

2004 Annual Book of ASTM Standards, Gaseous Fuels; Coal and Coke, vol. 05.06. ASTM, Philadelphia, PA, pp. 504-507.
Bianchini M, Frasconi P, Gori M (1995). Learning without local minima in radial basis function networks. IEEE Trans. Neural Networks1. 6(3): 749-755.
Broomhead DS, Lowe D (1988). Multivariable functional interpolation and adaptive networks. Complex Syst. 2: 321–355.
Chen S, Cowan CFN, Grant PM (1991). Orthogonal least squares learning algorithm for radial basis function networks. IEEE Trans. Neural Networks. 2(2): 302-309.
Finol J, Guo YK, Jing XD (2001). A rule based fuzzy model for the prediction of petrophysical rock parameters. L. Petr. Sci. Eng. 29: 97-113.
Foody GM (2004). Supervised image classification by MLP and RBF neural networks with and without an exhaustively defined set of classes. Int. J. Remote Sensing 25(15): 3091-3104.
Harpham C, Dawson CW (2006). The effect of different basis functions on a radial basis function network for time series prediction: A comparative study. Neurocomputing. 69: 2161-2170.
Jin Y, Jiang J (1999). Techniques in neural-network based fuzzy system identification and their application to control of complex systems. In: C.T. Leondes, Editor, Fuzzy Theory Systems, Techniques and Applications, Academic Press, New York, pp. 112-128.
Kaynar O, Yilmaz I, Demirkoparan F (2011). Forecasting of natural gas consumption with neural network and neuro fuzzy system. Energy Education Science and Technology Part A: Energy Sci. Res. 26: 221-238.
Marschalko M, Fuka M, Treslin L (2008). Influence of mining activity on selected landslide in the Ostrava-Karvina coalfield. ACTA Montanistica Slovaca. 13: 58-65.
Marschalko M, Treslin L (2009). Impact of underground mining to slope deformation genesis at Doubrava Ujala. ACTA Montanistica Slovaca. 14: 232-240.
Matlab 7.1 (2005). Software for technical computing and model-based design. The MathWorks Inc.
Moody J, Darken CJ (1989). Fast learning in networks of locally-tunes processing units. Neural Computation 1(2): 281-294.
Park J, Sandberg IW (1991). Universal approximations using Radial-Basis-Function Network. Neural Computation. 3(2): 246-257.
Park J, Sandberg IW (1993). Approximation and Radial Basis Function Networks Neural Computation. 5: 305-316.
Rivas VM, Merelo JJ, Castillo PA, Arenas MG, Castellano JG (2004). Evolving RBF neural networks for time-series forecasting with EvRBF. Info. Sci. 165: 207-220.
Rumelhart D, McClelland J (1986). Parallel distributed processing: explorations in the microstructure of cognition. Bradford books, MIT Press, Cambridge.
Sarimveis H, Doganis P, Alexandridis AA (2006). Classification technique based on radial basis function neural Networks. Adv. Eng. Software. 37: 218–221.
Sheta AF, Jong KD (2001). Time-series forecasting using GA-tuned radial basis function. Info. Sci. 133: 221-228.
Singh TN, Kanchan R, Verma AK, Singh S (2003). An intelligent approach for prediction of triaxial properties using unconfined uniaxial strength. Min. Eng. J. 5: 12–16.
SPSS 10.0.1. (1999). Statistical analysis software (Standard Version). SPSS Inc.
Turkish Lignite Authority (TKI) web page (2008). www.tki.gov.tr (in Turkish).
Xu K, Xie M, Tang LC, Ho SL (2003). Application of neural networks in forecasting engine systems reliability. Appl. Soft Computing. 2: 255-268.
Yilmaz I (2006). Indirect estimation of the swelling percent and a new classification of soils depending on liquid limit and cation exchange capacity. Eng. Geol. 85: 295-301.
Yilmaz I (2009a). Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: a case study from Kat landslides (Tokat-Turkey). Comput. Geosci. 35(6): 1125-1138.
Yilmaz I (2009b). Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: conditional probability, logistic regression, artificial neural networks, and support vector machine. Environmental Earth Sciences. DOI: 10.1007/s12665-009-0394-9.
Yilmaz I (2010). The effect of the sampling strategies on the landslide susceptibility mapping by Conditional Probability (CP) and Artificial Neural Networks (ANN). Environ. Earth Sci. 60(3): 505-519.
Yilmaz I, Yüksek AG (2008). An example of artificial neural network application for indirect estimation of rock parameters. Rock Mech. Rock Eng. 41: 781-795.
Yilmaz I, Yüksek AG (2009). Prediction of the strength and elasticity modulus of gypsum using multiple regression, ANN, ANFIS models and their comparison. Int. J. Rock Mech. Min. Sci. 46(4): 803-810.
Yu L, Lai KK, Wang S (2008). Multistage RBF neural network ensemble learning for exchange rates forecasting. Neurocomputing 71: 3295-3302.
Zhang R, Huang G, Sundararajan N, Saratchandran P (2007). Improved GAP-RBF network for classification problems. Neurocomputing. 70: 3011-3018.