J. Basic. Appl. Sci. Res., 3(6)892-898, 2013
© 2013, TextRoad Publication

ISSN 2090-4304
Journal of Basic and Applied Scientific Research
www.textroad.com

Benchmarking by an Integrated Data Envelopment Analysis-Artificial Neural Network Algorithm

Leila Karamali[1*], Azizollah Memariani[2], Gholam Reza Jahanshahloo[3], Mohsen Rostamy-malkhalifeh[4]

[1,3,4]Department of Mathematics, Science and Research Branch, Islamic Azad University, Tehran, Iran
[2]Department of Mathematics and Computer Science, University of Economic Sciences, Tehran, Iran

ABSTRACT

Data envelopment analysis (DEA) is a nonparametric approach using mathematical models, which evaluates the efficiency in a set of decision making units (DMUs) and offers the benchmarks to the inefficient units for better performance. Artificial neural networks (ANNs) are configured for specific applications, such as pattern recognition, function approximation, data classification and so on in different areas of sciences. In this paper, an algorithm is proposed using DEA and ANN for efficiency analysis and benchmarking. One of the important issues, from the managers' point of view, is to improve the efficiency of the DMUs by altering a given parameter and subsequently finding appropriate benchmark for this DMU. In the four-stage proposed algorithm, first the efficient units are identified by DEA, then the coordination of inputs and outputs related to the efficient DMUs are used for training the ANN in order to establish a correlation among these entities. Managers' desired inputs are given to the trained ANN, so the outputs are estimated for future. The new set of input-output coordination is applied to DEA in order to analyze the performance and obtaining benchmark to the inefficient DMUs. The proposed algorithm has been incorporated in a banking system. The results of this algorithm provides useful information on the evaluation of DMUs' efficiency and also benchmarking for the inefficient DMUs, for future periods based on the managers' desired input values.

KEYWORDS: Data Envelopment Analysis (DEA), Artificial Neural Network (ANN), Benchmarking.

1. INTRODUCTION

Management and control requires the evaluation and prediction of the DMUs in efficiency analysis settings. Among the existing methods, DEA and ANN have proved to be most appropriate tools for this kind of analysis. One of the important problems of DEA is that, it evaluates the efficiency in the past and is not capable to forecast the efficiency frontier for future, while it is necessary in order to get a good decision to improve the inefficient DMUs' performance. Therefore, ANN and DEA are integrated to solve this problem and the obtained results avail the managers to construct an effective system for forecasting a suitable benchmark to the inefficient DMUs.

There had been a numerous works incorporating DEA and ANN for efficiency analysis in the literatures:

On performance and efficiency analysis: Athanassopoulos and Curram [1] concerned the integration of DEA and ANN as tools for assessing performance. They compared both methods in a practical case on a set of bank branches, having multi-inputs and multi-outputs. After selecting data, the ANNs trained in learning a model for nonlinear forecasting. Although, there are differences between both methods, but their results have been demonstrated that, a useful range of information could be offered to the assessment of performance by both models. Costa and Markellos [2] showed how the results of ANN are similar to DEA. Pendharkar and Rodger [3] utilized DEA to screen databased approaches in constructing a sub-sample of training data, which is almost monotonic, and in certain forecasting problems, it is a basic property assumed. Obtained results showed the forecasting ability of ANNs acquired by training on efficient data set, is stronger than the performance that is forecasted by an ANN, which is trained from the set of inefficient one. Wang [4] proposed a non-parametric efficiency analysis based on the adaptive neural networks. Wu et al. [5] integrated data envelopment analysis and neural networks to examine the relative branch efficiency of a big Canadian bank and compared the results with DEA results. Celebi and Bayraktar [6] explored a novel integration of ANN and DEA for evaluation of suppliers under incomplete information of evaluation criteria. Wu [7] presented a hybrid model using data envelopment analysis, decision trees and artificial neural networks to assess supplier performance. Azadeh et al. [8] proposed an algorithm that assesses the impact of personnel efficiency attributes on total efficiency via DEA and ANN. DEA provides data for ANN and ANN results

[*]**Corresponding Author:** Leila Karamali, Department of Mathematics, Science and Research Branch, Islamic Azad University, Tehran, Iran. Email: lm_karamali@yahoo.com

is the basis for forecasting total efficiency. Azadeh et al. [9] introduced a non-parametric efficiency frontier analysis method on adaptive neural network.

On classification and clustering: Mohamed M. Mostafa [10] used DEA and artificial neural network to model and classify the relative efficiency of top Arab banks. Samoilenko et al. [11] augmented DEA with cluster analysis (CA) and artificial neural networks (ANNs) and proposed a five-step methodology to investigate whether the difference in the scores of scale heterogeneous DMUs is due to the heterogeneity of the levels of inputs and outputs, or whether it is caused by their efficiency of mapping of inputs into outputs. Pendharkar [12] suggested a hybrid radial basis function network- DEA neural network for classification problems. DEA is used to learn the classification function.

On frontier analysis: Santin et al. [13] used ANN in different observed cases to simulate nonlinear production function for comparing its performance with standard methods such as DEA and stochastic frontier. The outcome of their suggested method showed that in conventional approaches, ANN is a promising method in fitting production functions and measuring efficiency under nonlinear assumptions

On other fields: Azadeh et al. [14] studied the capability of handling complexity and nonlinearity through DEA and ANN. Kheirkhah et al. [15] worked on the impact of data preprocessing and post processing on artificial neural network performance. They utilized DEA to compare constructed artificial neural network models.

So far, we have not noticed an attempt for estimating the efficiency of DMUs based on the desired alteration of parameters by decision maker and finding benchmark based on this alteration, in the literature. Inability of DEA in forecasting is the reason of proposing the algorithm in this paper. In the proposed algorithm, DEA plays two various roles. First, it prepares the data required for ANN, second it uses coordination of inputs and outputs that are produced by ANN to provide the suitable benchmark for inefficient DMUs. ANN uses the inputs and outputs of the efficient DMUs in successive periods, respectively, as the inputs and target vectors in training process. The proposed algorithm has been implemented in a banking system.

In section 2, data envelopment analysis and artificial neural network are introduced. Section 3 proposes the integrated algorithm. Section 4 provides the case study of the integrated algorithm. Finally, in section 5 a compendious conclusion is presented.

2. Data Envelopment Analysis and Artificial Neural Network

2.1 Data Envelopment Analysis (DEA)

Since in practical problems, especially in banking systems, the return to scale is variable, in this context we use BCC model. The input-oriented BCC model with variable returns to scale (1) evaluates the relative efficiencies of *n* DMUs *(j=1,…,n)*, each with *m* inputs and *s* outputs denoted by $x_{1j}, x_{2j}, \ldots, x_{mj}$ and, $y_{1j}, y_{2j}, \ldots, y_{sj}$ respectively, by minimizing inputs when outputs are constant. Model (1) in the input-oriented BCC, is as follows:

$$\begin{aligned}
& Min \quad \theta \\
& s.t \quad \sum_{j=1}^{n} \lambda_j x_{ij} \le \theta x_{io}, \qquad i=1,\ldots,m, \\
& \qquad \sum_{j=1}^{n} \lambda_j y_{rj} \ge y_{ro}, \qquad r=1,\ldots,s, \\
& \qquad \sum_{j=1}^{n} \lambda_j = 1 \\
& \qquad \lambda_j \ge 0, \qquad\qquad j=1,\ldots,n.
\end{aligned} \tag{1}$$

The calculations provide an efficiency score θ, using linear optimization for each *DMU* with respect to the closest observation on the frontier based on orientation of the model. Also the variable λ_j score is helpful in determining the benchmark for inefficient DMUs. If $\lambda_j = 0$, then DMU_j doesn't play any role in determining the benchmark and if $\lambda_j > 0$ then DMU_j is effective in considered benchmark as the coefficient λ_j .

2.2 Artificial Neural Networks

Artificial Neural Network (ANN) incorporates to correlate the input and output parameters. Input parameters of the DMUs are taken as the input layer and the output parameters as the output layer of the network. The network learning is done with the data available in different periods for the efficient DMUs. This is to find the correlation of the input and output parameters when the DMUs are efficient.
From the existing learning rules, Backpropagation (BP) that includes searching weights by minimizing the cost function is utilized. BP learning rule is a supervised learning algorithm.

3. The proposed algorithm

DEA is well known for evaluating the efficiency of DMUs and benchmarking for inefficient DMUs, but it fails to obtain the efficiency and benchmark for future based on the desired value of the input parameters by decision maker. This information is in demand for most managers. Therefore, a complementary tool is to be utilized along with the DEA to perform this part of the analysis. The purpose of the proposed algorithm in this study is to provide benchmark offering to the managers for inefficient units, using ANN when the information for the input parameters is altered. A similar network can be developed when the inputs are to be forecasted by altering the output parameters. For which the output parameters will be considered as the input layer and input parameters as the output layer of the network.

Steps of the proposed algorithm are portrayed in Figure 1. Details of the proposed algorithm will be discussed in the following four steps.

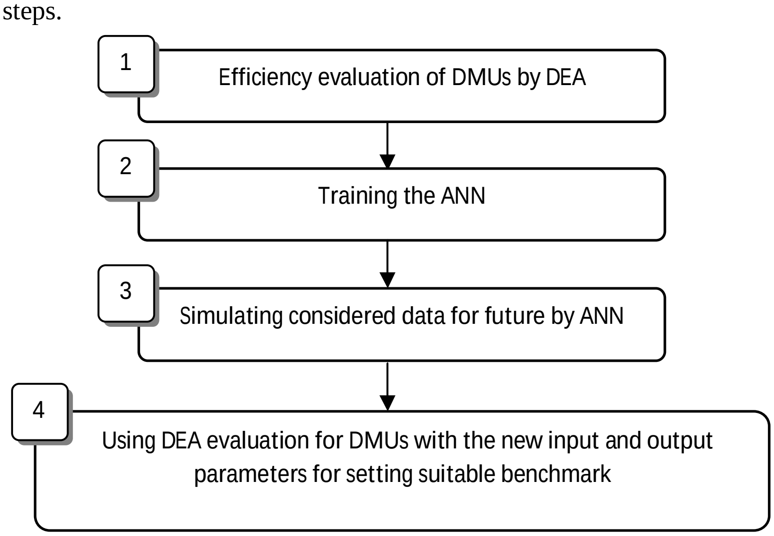

Figure1. An integrated DEA-ANN Algorithm for forecasting efficiency and benchmarking

3.1 Step 1: Efficiency evaluation of DMUs by DEA

DEA is a multivariate analysis, which evaluates efficiency of a set of DMUs and considers many inputs and outputs in efficiency assessment. As by different quantity of inputs and outputs the efficiency scores may vary, thus it is needed to consider accurate specifications of DEA in each case. In this stage efficiency of DMUs, in every periods, will be evaluated (by model 1) with input oriented BCC model. As the frontier of production function is constructed by efficient DMUs (best practices), we want to estimate the production function via training ANN. In this step, we specify efficient units in all periods, so the input-output specifications of these efficient DMUs will be referred to the next step.

3.2 Step 2: Training the ANN

In this step the information of input and output parameters for the efficient DMUs in all periods are collected. An ANN is trained with the values of input parameters as the input layer and the value of output parameters as the output layer. Training step will be performed by Resilient Backpropagation (Rprop) training algorithm. In this method, the data set is divided into three subsets: training, validation and test. Training set is used for computing the gradient and updating the network weights and biases. The error on the validation set is monitored during the training process. The validation error will normally decrease during the initial phase of training, as does the training set error. When the network begins to over fit the data, the error on the validation set will typically begin to rise. When the validation error increases for a specified number of iterations, the training is stopped and the weights and biases at the minimum of the validation error are returned [16]. The test set error is not used during training, but it is used to compare different models.

3.3 Step 3: Simulating considered data for future by ANN

In this step, decision maker executes the trained ANN for the desired input of efficient DMUs in order to predict the efficiency frontier. The output of the network will be the estimated output parameters.

3.4 Step 4: Using DEA evaluation for DMUs with the new input and output parameters for setting suitable benchmark

Finally, DEA model is run for the new set of information and the efficiency score of the inefficient DMUs will be predicted. If they remain inefficient then respected benchmarks will be determined.

4. The case study

The combined proposed algorithm has been used for a large bank in 28 successive periods. The case study on 338 branches, for efficiency evaluation and setting suitable benchmark for inefficient DMUs, in the last period has been considered. The procedural application of the algorithm will be detailed in the following:

4.1 Efficiency evaluation of DMUs by DEA

Here inputs and outputs of each bank branches are considered as the same inputs and outputs in DEA model as well as ANN architecture. Table 1 shows a sample of inputs and outputs used in DEA and ANN.

Table1. DEA (ANN) inputs and outputs

DEA (ANN) inputs	DEA (ANN) outputs
Personnel scores	Facilities
Paid profits	Sum of deposits
Pending demands	Received profits
	Received wages
	Other resources

In this step DEA model (1) is run for all 338 DMUs in each period and after specifying all efficient DMUs their input-output specifications are to be sent in to the next step.

4.2 Training the ANN

The preferred ANN in this application is a two- layer neural network with a single hidden layer. In both of the layers, Tansig transfer function has been used. In addition, for training the network, Trainrp algorithm (Resilient Backpropagation (Rprop) training algorithm) has been utilized. Inputs and outputs of the network are respectively, the input and output parameters of DEA, thus the input layer is a vector with three elements and the output layer has five neurons. The parameter q indicates the number of neurons in hidden layer, which is 21 as best fitted. (Figure2).

Error in the network has been calculated by msereg[1]. For improving the generalization of the ANN, we used early stopping method. In this method, data set is divided into three subsets: training 80%, validation 10% and test 10% of whole data set.

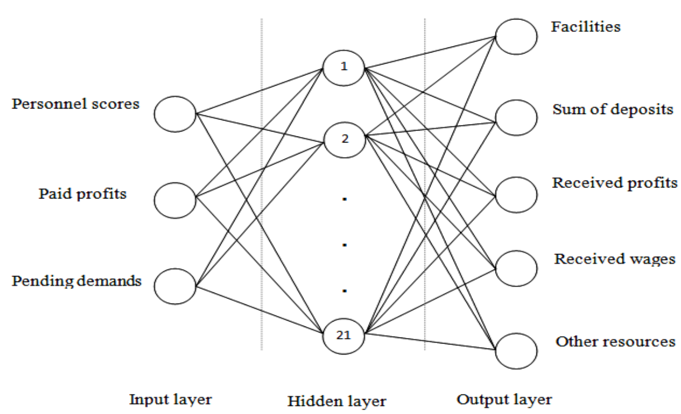

Figure2. Topology of the proposed network

[1] Mean Squared Error with Regularization

4.3 Simulating considered data for future by ANN

In this step, the inputs and outputs of those DMUs evaluated as efficient in DEA in 28 periods, for 338 bank branches that are totally 1075 are at hand for training process by ANN. DEA inputs are considered as inputs of the network and the same outputs of DEA are considered as outputs of the network. Also by using Tansig transfer functions for both layers and the learning rule as mentioned earlier for determining the value of q. This process will be continued until correlation between outputs and the targets reaches to 1 as much as possible. After performing these processes, suitable quantity for q has been obtained as 21. Figure 3 shows performance value as 6.63e+22 in training processes. The correlation coefficient in training, validation and test process between outputs and targets are illustrated in figure 4. In figure 5 the correlation coefficients for each output are illustrated.

- Specification of the inputs and outputs of efficient units in the final period, respectively, replaced by the inputs considered from the manager's point of view and those outputs simulated by the mentioned network.
- In addition, specifications of the inefficient units in inputs considered from manager too, and the outputs are the same as final period.

Now the required data for DEA is at hand.

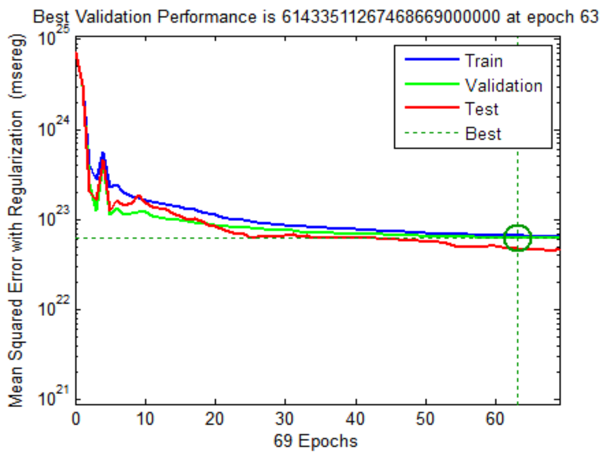

Figure3. Error in the training, validation and test process

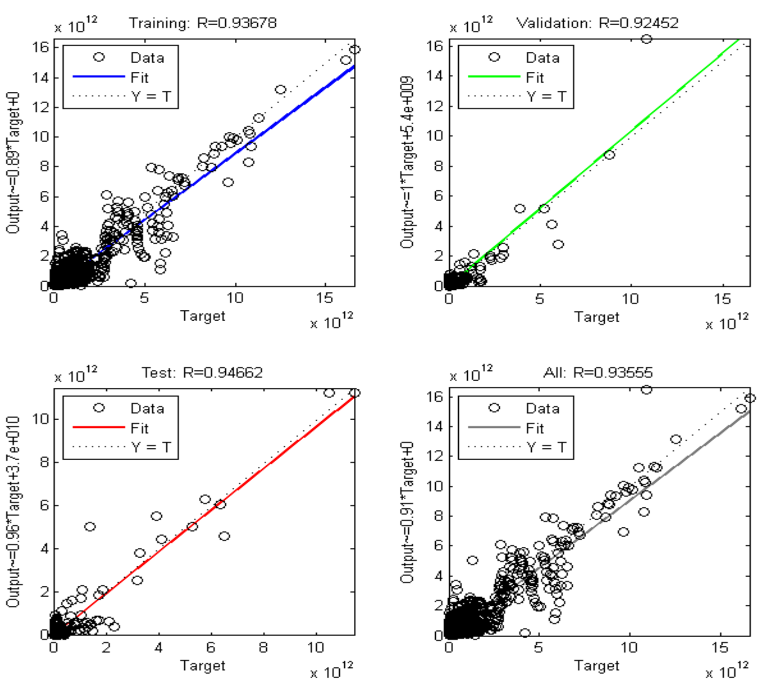

Figure4. The R-value in training, validation and test set

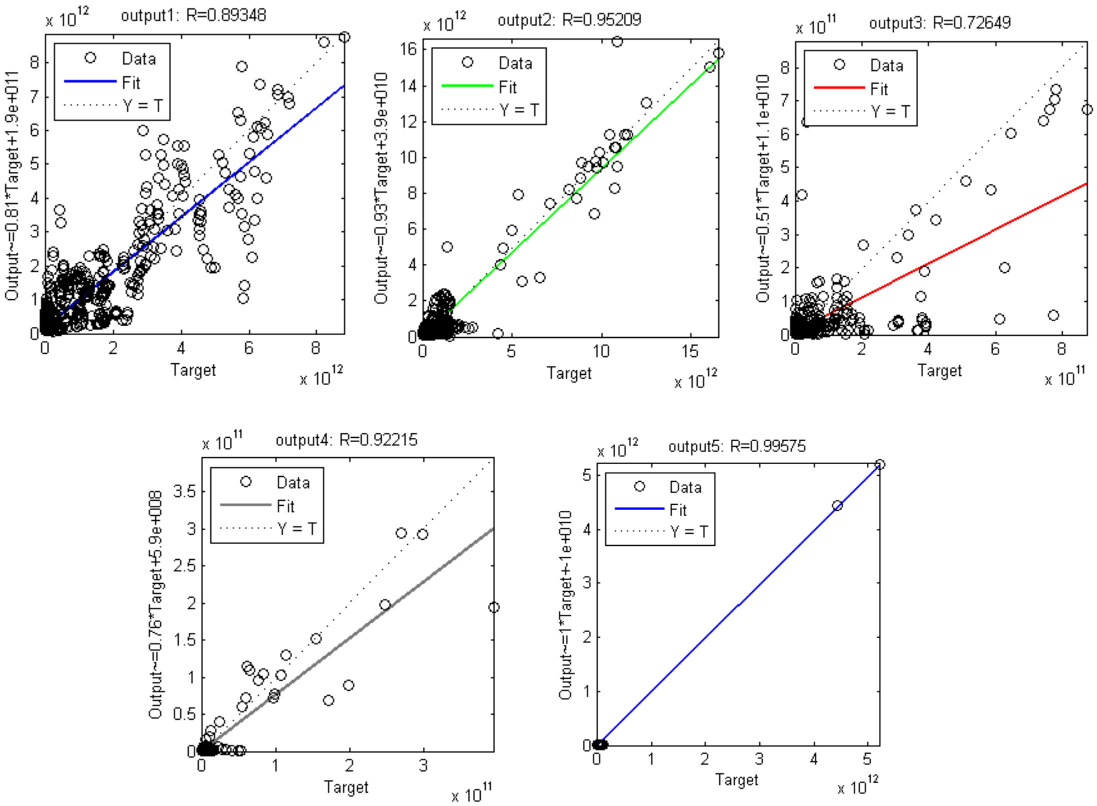

Figure5. The R-value between outputs and targets in the process of training

4.4 Using DEA evaluation for DMUs with the new input and output parameters for setting suitable benchmark

After performing DEA model, the efficiency of all units and benchmark for each inefficient unit, will be acquired. Therefore, according to the mentioned method it is possible to set benchmark for each inefficient unit for future. Results show that 75 percent of efficient DMUs remained efficient. In order to survey the efficiency deviations of efficient DMUs in the last period compared with the efficiencies obtained in step 4, MAE[2] method is used. The average variations in efficiency scores of efficient DMUs are obtained as 0.05, which shows a rational deviation. Also by this forecasting, some of inefficient DMUs with efficiency close to 1 became efficient. Finally, the number of efficient DMUs in forecasting results increased from 44 to 49 DMUs. Increasing the number of efficient DMUs shows performance improvement of the system.

Benchmark inputs for 10 inefficient units in final period are listed in Table2. It should be noted that since we used BCC (input-oriented) model, benchmark outputs do not change any more compared to the actual outputs for the inefficient DMUs. So in table 2 we did not consider the outputs.

Table2. Some of inefficient DMUs and their future benchmarks by the input oriented BCC model

Some of Inefficient DMUs		**Real inputs**			**Benchmark inputs**		
	Efficiency	**I_1**	**I_2**	**I_3**	**I_1**	**I_2**	**I_3**
DMU_7	0.9553	4.75	703004412	913734436	4.711839	6.34E+08	8.65E+08
DMU_{41}	0.9069	9.07	940639738	1.846E+09	5.737759	5.41E+08	1.11E+09
DMU_{43}	0.7161	7.45	573588190	118585966	6.158899	9.28E+08	2.2E+08
DMU_{55}	0.7389	9.47	2.34E+09	1.713E+10	7.987485	1.79E+09	1.38E+10
DMU_{108}	0.3153	17.3	3.407E+09	3.177E+09	4.662896	8.35E+08	8.17E+08
DMU_{170}	0.8588	21.44	2.348E+10	4.349E+10	16.43622	1.64E+10	3.18E+10
DMU_{196}	0.6872	8.28	969733321	309028577	6.180731	6.58E+08	2.2E+08
DMU_{266}	0.2639	23.35	3.297E+09	4.733E+10	6.007732	7.71E+08	1.16E+10
DMU_{275}	0.4895	12.35	4.556E+09	1.443E+09	6.223289	2.09E+09	6.94E+08
DMU_{337}	0.5537	38.94	4.54E+09	7.687E+10	29.88357	3.17E+09	5.63E+10

[2] Mean Absolute Error

As can be seen in table 2, benchmark inputs are less than real inputs. It means each of inefficient DMUs used more inputs than it needed, so they are inefficient. The result of table 2 explains each inefficient DMU can be efficient if it uses fewer inputs as the same as benchmark inputs.

5. CONCLUSION

The integrated approach facilitates the managers for performing what if analysis by modifying the input values of the inefficient DMUs. The input values are allowed to be altered and the output values are estimated. This estimation is based on the behavior of efficient DMUs how they correlate the input and output parameters. The new set of information is fed to DEA model again to check the change in the efficiency score of inefficient units and if they remain inefficient then appropriate benchmark is determined. In any system, when the manager, frequently like to simulate the score of inefficient benchmarks and try to push them towards efficiency, this model provides useful information to them.

For future researches, sensitivity analysis can be studied on the variation of input and output parameters by ANN and the effect of these variations on other variables in DEA.

REFERENCES

1. Athanassopoulos, A.D. and S.P. Curram (1996). A comparison of data envelopment analysis and artificial neural networks as tool for assessing the efficiency of decision making units. *Journal of Operational Research Society*, 47, 1000-1016.
2. Costa A. and R.N. Markellos (1997). Evaluating Public Transport Efficiency with Neural Network Models. Transportation Research Part C: *Emerging Technology*, 5, 301-312
3. Pendharkar, P.C. and J.A. Rodger (2003). Technical efficiency-based selection of learning cases to improve forecasting accuracy of neural networks under monotonicity assumption. *Decision Support Systems*, 36, 117-136.
4. Wang, S. (2003). Adaptive non-parametric efficiency frontier analysis: a neural-network-based model Original Research Article. *Computers & Operations Research*, 30(2), 279-295.
5. Wu, D., Z. Yang and L. Liang (2006). Using DEA-neural network approach to evaluate branch efficiency of a large Canadian bank. *Expert Systems with Applications*, 31, 108-115.
6. Çelebi, D. and D. Bayraktar (2008). An integrated neural network and data envelopment analysis for supplier evaluation under incomplete information. *Expert Systems with Applications*, 35, 1698-1710.
7. Wu, D. (2009). Supplier selection: A hybrid model using DEA, decision tree and neural network. *Expert Systems with Applications*, 36 (5), 9105-9112.
8. Azadeh, A., M. Saberi, R. Tavakkoli Moghaddam and L. Javanmardi (2011). An integrated Data Envelopment Analysis–Artificial Neural Network–Rough Set Algorithm for assessment of personnel efficiency. *Expert Systems with Applications*, 38 (3), 1364-1373.
9. Azadeh, A., M. Saberi, M. Anvari, A. Azaron and M. Mohammadi (2011). An adaptive network based fuzzy inference system–genetic algorithm clustering ensemble algorithm for performance assessment and improvement of conventional power plants. *Expert Systems with Applications*, 38(3), 2224-2234.
10. Mostafa, M. M. (2009). Modeling the efficiency of top Arab banks: A DEA–neural network approach. *Expert Systems with Applications*, 36 (1), 309-320.
11. Samoilenko, S. and K. M. Osei-Bryson (2010). Determining sources of relative inefficiency in heterogeneous samples: Methodology using Cluster Analysis, DEA and Neural Networks. *European Journal of Operational Research*, 206 (2), 479-487.
12. Pendharkar, P.C. (2011). A hybrid radial basis function and data envelopment analysis neural network for classification. *Computers & Operations Research*, 38(1), 256-266.
13. Santin, D., F.J. Delgado and A.Valino (2004). The measurement of technical efficiency: A neural network approach. *Applied Economics*, 36, 627-635.
14. Azadeh, A., A. M. Sheikhalishahi and S. M. Asadzadeh (2011). A flexible neural network-fuzzy data envelopment analysis approach for location optimization of solar plants with uncertainty and complexity. *Renewable Energy*, 36 (12), 3394-3401.
15. Kheirkhah, A., A. Azadeh, M. Saberi, A. Azaron and H. Shakouri (2012). Improved Estimation of Electricity Demand Function by Using of Artificial Neural Network, Principal Component Analysis and Data Envelopment Analysis. *Computers & Industrial Engineering, In Press, Accepted Manuscript, Available online 13 October 2012.*
16. Azadeh A., M. Saberi, M. Anvari (2010) An integrated artificial neural network algorithm for performance assessment and optimization of decision making units. *Expert Systems with Applications,* 37, 5688-5697.