Expert Systems with Applications 38 (2011) 1364–1373

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

An integrated Data Envelopment Analysis–Artificial Neural Network–Rough Set Algorithm for assessment of personnel efficiency

Ali Azadeh [a,*], Morteza Saberi [b,c], Reza Tavakkoli Moghaddam [a], Leili Javanmardi [a]

[a] Department of Industrial Engineering and Center of Excellence for Intelligent-Based Experimental Mechanics, College of Engineering, University of Tehran, Iran
[b] Department of Industrial Engineering, University of Tafresh, Iran
[c] Institute for Digital Ecosystems & Business Intelligence, Curtin University of Technology, Perth, Australia

ARTICLE INFO

Keywords:
Personnel
Efficiency
Data Envelopment Analysis (DEA)
Artificial Neural Network (ANN)
Rough Set Theory (RST)
Cross Validation Test Technique (CVTT)

ABSTRACT

Personnel specifications have greatest impact on total efficiency. They can help us to design work environment and enhance total efficiency. Determination of critical personnel attributes is a useful procedure to overcome complication associated with multiple inputs and outputs. The proposed algorithm assesses the impact of personnel efficiency attributes on total efficiency through Data Envelopment Analysis (DEA), Artificial Neural Network (ANN) and Rough Set Theory (RST). DEA has two roles in the proposed integrated algorithm of this study. It provides data ANN and finally it selects the best reduct through ANN result. Reduct is described as a minimum subset of attributes, completely discriminating all objects in a data set. The reduct selection is achieved by RST. ANN has two roles in the integrated algorithm. ANN results are basis for selecting the best reduct and it is also used for forecasting total efficiency. The proposed integrated approach is applied to an actual banking system and its superiorities and advantages are discussed.

0. Significance

This is the first study which proposes an integrated algorithm for assessment of the impact of personnel efficiency attributes on total efficiency through Data Envelopment Analysis (DEA), Artificial Neural Network (ANN) and Rough Set Theory (RST). The outcome helps managers to construct helpful system to forecast efficiencies by selected attributes. The integrated algorithm is successfully applied to 102 branches of a large private bank, evaluating personnel attributes impact on bank branches efficiencies. The results of this study show that four subsets of conditional attributes with total number of nine attributes from 28 attributes have a critical impact on the accuracy of the optimal solution. This reduction in attributes number decrease the time of decision-making and consequently reduces the cost of efficiency evaluation.

1. Introduction

Efficiency is a key concept for organizations. Too many immeasurable influences and complex relationships among attributes influence efficiency in organizations. Efficiency relevant to human attributes is a goal that is rarely questioned in contemporary organizations. As personnel specifications have greatest impact on efficiency, they can help us to design work environments and enhance total efficiency. As providing information on multiple input and output factors are a complicated and time-consuming procedure, determining critical personnel attributes is useful. The purpose of proposed integrated algorithm in present study is to alert management to the important attributes that should be considered if an effective decision to enhance efficiency is to be formulated. This is because there is a great desire to identify the critical attributes for sensitivity analysis of inefficient decision-making units (DMUs) regarding efficiency attributes. The algorithm proposes an analytic function that predicts these attributes exactly. This model is applicable for all problems associated with decision-making in organizations composed of decision-making units (DMUs) and will be valuable for executives and senior managers. The outcome helps managers to construct helpful system to forecast DMUs efficiency by selected attributes. Furthermore, reduction in attributes number decrease the time of decision-making and consequently reduces the cost of efficiency analysis.

The integrated algorithm uses Data Envelopment Analysis (DEA), ANN and Rough Set Theory (RST). DEA has two roles in the proposed integrated algorithm of this study. It provides data ANN and finally it selects the best reduct. Reduct is described as a minimum subset of attributes, completely discriminating all objects in a data set. The reduct selection is achieved by RST. ANN has two roles in the integrated algorithm. ANN results are basis for

* Corresponding author.
E-mail address: aazadeh@ut.ac.ir (A. Azadeh).

0957-4174/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2010.07.033

selecting the best reduct and it is used for forecasting total efficiency. The proposed algorithm is applied to an actual banking system.

Weakness of DEA in forecasting is the reason to use ANN (Wu, Yang, & Liang, 2006). ANN has been viewed as a good tool to approximate nonlinear problems and is useful for manager in predicting system (Cybenko, 1989; Patuwo, Hu, & Hung, 1993). In addition, RST and ANN are combined to construct the forecasting feature of the proposed algorithm. The best reduct is selected when ANN performance is at its best with the selected reduct. In order to apply necessity available error measurement of ANN in best reduct determination procedure, DEA is used.

1.1. ANN and efficiency

Athanassopoulos and Curram (1996) first introduced the idea of combination of neural networks and DEA for classification and/or prediction. They treated DEA as a preprocessing methodology to screen training cases in a study. Their application is bank with multi-output: 4 inputs, 3 outputs. After selecting samples, the ANNs are then trained as tools to learn a nonlinear forecasting model. They assume that inefficiency distributions are semi-normal and exponential and conclude that DEA is superior to ANN for measurement purpose. Their study indicates that ANN results are more similar with the constant returns to scale and less with the variable returns to scale results. The latter, is a consequence of the implicit assumption of constant returns to scale adopted by the ANN models.

Costa and Markellos (1997) analysed the London underground efficiency with time series data for 1970–1994 where there are 2 inputs – fleet and workers – and 1 output – km s. They explain how the ANNs results are similar to COLS and DEA. They proposed two procedures: (a) similar way to COLS after neural training; (b) by an oversized network until some signal to noise ratio is reached. Then, inefficiency is determined as observation-frontier distance. However, ANNs offer advantages in the decision-making, the impact of constant versus variable returns to scale or congestion areas (Costa & Markellos, 1997). Santin and Valino (2000) study on education efficiency by a two-level model: student – production function is estimated by ANNs – and school. They infer that ANN is superior to econometric approach at frontier estimation. Pendharkar and Rodger (2003) used DEA as a data screening approach to create a sub sample training data set that is 'approximately' monotonic, which is a key property assumed in certain forecasting problems. Their results indicate that the predictive power of an ANN trained on the 'efficient' training data subset is stronger than the predictive performance of an ANN trained on the 'inefficient' training data subset. Santin, Delgado, and Valino (2004) used a neural network approach for a simulated nonlinear production function and compared its performance with conventional alternatives such as stochastic frontier and DEA in different observations and noise scenarios. The results suggested that ANNs are a promising alternative to conventional approaches, to fit production functions and measure efficiency under nonlinear contexts. Wu et al. (2006) presented a DEA–NN study for performance assessment of branches of a large Canadian bank. The results are operable to the normal DEA results overall. They concluded that the DEA–NN approach produces a more robust frontier and identifies more units that are efficient because better performance patterns are explored. Furthermore, for worse performers, it provides the guidance on how to improve their performance to different efficiency ratings. Ultimately, they concluded the neural network approach requires no assumptions about the production function (the major drawback of the parametric approach) and it is highly flexible. ANNs have been viewed as a good tool to approximate numerous nonparametric and nonlinear problems.

Azadeh, Ghaderi, Anvari, and Saberi (2006a), Azadeh, Ghaderi, Anvari, Saberi, and Izadbakhsh (2007) and Azadeh, Ghaderi, Anvari, and Saberi (2007) proposed a highly unique flexible ANN algorithm to measure and rank the decision-making unit's (DMUs) efficiency. Their algorithm calculated efficiency score of Iran steam power plants in 2004. Results indicate that the proposed algorithm estimates the values of efficiency scores closer to the ideal efficiency. They concluded that the propose algorithm estimates more robust results and more efficient units than the conventional approach because better performance patterns are explored. In addition, they proposed a method to integrate their pervious algorithm (Azadeh, Anvari, & Saberi, 2007, 2008).

1.2. Rough Set Theory

Several immeasurable influences and complex relationships among attributes impact efficiency in organizations. Rough Set Theory (RST) proposed by Pawlak, is one of the techniques for the identification and recognition of common patterns in data Pawlak (1982, 1991). This technique has found applications in knowledge discovery from data bases, data mining, fault diagnosis, machine learning, knowledge acquisition, expert systems and decision support systems (Błaszczyński, Greco, & Słowiński, 2007; Fan, Liu, & Tzeng, 2007; Inuiguchi & Miyajima, 2007). It is also used to study uncertainty (Beynon & Peel, 2001; Lili & Zhi, 2001; Ziarko, 1993), prediction (Becerra-Fernandez, Zanakis, & Walczak, 2002; Kusiak, Kern, Kernstine, & Tseng, 2000; Sanchis, Segovia, Gil, Heras, & Vilar, 2007), service organizations (Chou, Cheng, & Chang, 2007; Hassanien, 2007; Kowalczyk & Slisser, 1997; Sikder & Gangopadhyay, 2007; Tsumoto, 1997), financial firms (Ravi Kumar & Ravi, 2007; Ruhe, 1996; Shyng, Wang, Tzeng, & Wu, 2007), and scheduling problems (Liu, Chen, Wu, & Li, 2006; Triantaphyllou, Liao, & Iyengar, 2002). Stefanowski and Slowinski have studied rough sets as a tool for feature selection by studying attribute dependencies (Stefanowski & Slowinski, 1997). Kusiak and Tseng (2000) have proposed two independent algorithms for accurate feature selection in medical, industrial and engineering case studies. Others like Xia and Wu discusses feature extraction technique of Rough Set Theory for supplier selection to select best suppliers according to different tangible and intangible attributes (Xia & Wu, 2007). Moreover, there are some other application of Rough Set Theory to feature selection in customer relationship management (Tseng & Huang, 2007), product quality evaluation (Zhai, Khoo, & Fok, 2002) and healthcare (Wang, Yang, Jensen, & Liu, 2006). However, existing heuristic rough set approaches to feature selection are insufficient at finding optimal reductions. On the other hand, it is not feasible to search for optimal in even in average sized datasets. Therefore, the combination of this method by other robust data mining tools may help practitioners to go further into feature selection to obtain more accurate results.

1.3. Data Envelopment Analysis

DEA is a non-parametric method that uses linear programming to calculate the efficiency in a given set of decision-making units (DMUs). The DMUs that make up a frontier envelop, the less efficient firms and the relative efficiency of the firms is calculated in terms of scores on a scale of 0 to 1, with the frontier firms receiving a score of 1. DEA models can be input or output oriented and can be specified as constant returns to scale (CRS) or variable returns to scale (VRS).

1.4. Artificial Neural Networks

An Artificial Neural Network (ANN) is an information-processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. Although ANNs arose to model the brain, they have been applied when there is no theoretical evidence about the functional form. In this way, ANNs are data-based, not model-based. The key element of this paradigm is the novel structure of the information-processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems. ANNs, like people, learn by example. An ANN is configured for a specific application, such as pattern recognition, function approximation, data classification and so on in different areas of science. ANNs are normally arranged in three layers of neurons, the so-called multilayer structures are input, hidden and output layers. Input layers, which are neurons (also called nodes or processing units), introduce the model inputs. Hidden layers combine the inputs with weights that are adapted during the learning process. Output layer provides the estimations of the network. Multi Layer Perceptron (MLP) is used as the basis of ANN in this study.

2. The integrated algorithm

Personnel specifications have greatest impact on organizational efficiency. They can help us design work environments and enhance total efficiency. As providing information on multiple input and output factors are a complicated and time-consuming procedure, determining critical personnel attributes is useful. The purpose of proposed integrated algorithm in present study is to alert management to the important attributes that should be considered if an effective decision to enhance total efficiency is to be formulated. The algorithm proposes an analytic function that predicts these attributes exactly.

There are two competing paradigms on efficiency analysis. Parametric and non-parametric approaches are widely used for the efficiency measurement. The first include the estimation of both deterministic and stochastic frontier functions (SFF) which is based on the econometric regression theory and has been widely accepted in the econometrics field. The latter include DEA and Free Disposal Hull (FDH), which are based on a mathematical programming approach. Each of these two methodologies has its strength as well as major limitations.

The non-parametric approach makes no assumption about the functional form of the frontier. Instead, it specifies certain assumptions about the underlying technology that in combination with the data set allows the construction of the production set. For instance, the DEA frontier is very sensitive to the presence of the outliers and statistical noise, which indicates that the frontier derived from DEA analysis, may be warped if the data are contaminated by statistical noise (Bauer, 1990). On the other hand, DEA can hardly be used to predict the performance of other decision-making units.

The integrated algorithm uses Data Envelopment Analysis (DEA), ANN and Rough Set Theory (RST). DEA has two roles in the proposed integrated algorithm of this study. It provides data for ANN and finally it selects the best reduct. Reduct is described as a minimum subset of attributes, completely discriminating all objects in a data set. The reduct selection is achieved by RST. ANN has two roles in the integrated algorithm. ANN results are basis for selecting the best reduct and it is also used for forecasting total efficiency. The proposed algorithm is applied to an actual banking system.

The weakness of DEA in forecasting is the reason to utilize ANN in our proposed algorithm (Wu et al., 2006). ANN has been viewed as an ideal tool to approximate nonlinear problems and is also a useful approach forecasting issues. This is why it is used as a part of our algorithm (Cybenko, 1989; Patuwo et al., 1993).

As numerous inputs are not useful for ANN modeling, Rough Set Theory and ANN are combined to resolve this issue. The number of inputs affects ANN in two ways. First, more inputs require larger network and this increases the risk of over fitting and increases the size of training set. Second, more inputs require longer time for network to converge to a set of weights. In addition, with too many inputs, the weights are less likely to be optimal. This is why we use RST in the proposed algorithm. Principle Component Analysis (PCA) and Factor Analysis (FA) could be also used for reduction in a data set. However, RST provides independent reducts, whereas, PCA or FA provide combined data. The best reduct is selected when ANN performance is at its best with the selected reduct. DEA is used to use available error measurement in reduct selecting procedure. The stages involved in the proposed algorithm are illustrated in Fig. 1. The components of the algorithm are presented in the following sections.

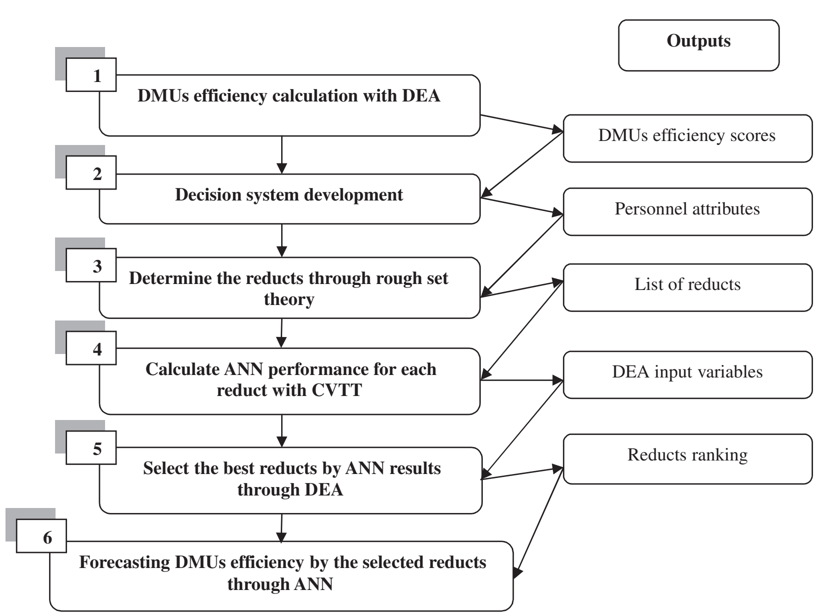

Fig. 1. An integrated DEA–ANN–Rough Set Theory Algorithm for decision-making units efficiency analysis and forecasting.

2.1. Step 1: DMUs efficiency calculation with DEA

This stage is involved with efficiency calculation of DMUs. DEA is a multivariate analysis tool that measures the relative efficiencies of a set of DMUs and its benefits are clearly understood (Azadeh, Amalnick, Ghaderi, & Asadzadeh, 2007; Azadeh, Ghaderi, & Izadbakhsh, 2007). It effectively considers multiple inputs and output factors in computing the efficiency scores. As efficiency scores vary on different selection of inputs and outputs, we should utilize an accurate DEA specification for each particular case (Serrano Cinca, Mar Molinero, & Chaparro García, 2002).

2.1.1. Basic models of DEA

The original fractional CCR model (1) evaluates the relative efficiencies of n DMUs ($j = 1 \cdots n$), 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 (Charnes, Cooper, & Rhodes, 1978). This is done so by maximizing the ratio of weighted sum of output to the weighted sum of inputs:

$$\begin{aligned}
&Max \quad \theta = \frac{\sum_{r=1}^{s} u_r y_{ro}}{\sum_{i=1}^{m} v_i x_{io}} \\
&s.t \quad \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1, \quad j = 1, \ldots, n, \quad r = 1, \ldots, s \\
&\qquad u_r, v_i \geq 0, \quad i = 1, \ldots, m \quad r = 1, \ldots, s
\end{aligned} \tag{1}$$

In model (1), the efficiency of DMU_o is θ_o and u_r and v_i are the factor weights. However, for computational convenience the fractional programming model (1) is re-expressed in linear program (LP) form as follows:

$$\begin{aligned}
&Max \quad \theta = \sum_{r=1}^{s} u_r y_{ro} \\
&s.t. \quad \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \leq 0, \quad j = 1, \ldots, n \\
&\qquad \sum_{i=1}^{m} v_i x_{io} = 1 \\
&\qquad u_r, v_i \geq \varepsilon, \quad i = 1, \ldots, m \quad r = 1, \ldots, s
\end{aligned} \tag{2}$$

where ε is a non-Archimedean infinitesimal introduced to ensure that all the factor weights will have positive values in the solution. The model (3) evaluates the relative efficiencies of n DMUs ($j = 1, \ldots, n$), respectively, by Minimizing inputs when outputs are constant. The dual of linear program (LP) model for input-oriented CCR is as follows:

$$\begin{aligned}
&Min \quad \theta \\
&s.t. \quad \theta x_{io} \geq \sum_{j=1}^{n} \lambda_j x_{ij}, \quad i = 1, \ldots m \\
&\qquad y_{ro} \leq \sum_{j=1}^{n} \lambda_j y_{rj} \quad r = 1, \ldots, s \\
&\qquad \lambda_j \geq 0
\end{aligned} \tag{3}$$

The output-oriented CCR model is as follows:

$$\begin{aligned}
&Max \quad \theta \\
&s.t. \quad x_{io} \geq \sum_{j=1}^{n} \lambda_j x_{ij}, \quad i = 1, \ldots m \\
&\qquad \theta y_{ro} \leq \sum_{j=1}^{n} \lambda_j y_{rj} \quad r = 1, \ldots, s \\
&\qquad \lambda_j \geq 0
\end{aligned} \tag{4}$$

If $\sum \lambda_j = 1$ ($j = 1, \ldots, n$) is added to model (3), the BCC model is obtained which is input oriented and its return to scale is variable. The calculations provide a maximal performance measure using piecewise linear optimization on each DMU with respect to the closest observation on the frontier. The linear programming system for the BCC input-oriented model is given in expression (5), and the output-oriented model in expression (6) (refer to Charnes, Cooper, Lewin, & Seiford (1994))

$$\begin{aligned}
&Min \quad \theta \\
&s.t. \quad \theta x_{io} \geq \sum_{j=1}^{n} \lambda_j x_{ij}, \quad i = 1, \ldots m \\
&\qquad y_{ro} \leq \sum_{j=1}^{n} \lambda_j y_{rj} \quad r = 1, \ldots, s \\
&\qquad \sum_{j=1}^{n} \lambda_j = 1 \\
&\qquad \lambda_j \geq 0, \quad j = 1, \ldots, n
\end{aligned} \tag{5}$$

$$\begin{aligned}
&Max \quad \theta \\
&s.t. \quad x_{io} \geq \sum_{j=1}^{n} \lambda_j x_{ij}, \quad i = 1, \ldots m \\
&\qquad \theta y_{ro} \leq \sum_{j=1}^{n} \lambda_j y_{rj} \quad r = 1, \ldots, s \\
&\qquad \sum_{j=1}^{n} \lambda_j = 1 \\
&\qquad \lambda_j \geq 0, \quad j = 1, \ldots, n
\end{aligned} \tag{6}$$

2.2. Step 2: Decision system definition

A data set or information system is a table, where each row indicates an object and each column shows an attribute that can be measured for each object. The input attributes are called conditional and the output is called decision attribute, respectively. In this study, the information system is comprised of decision-making units (DMUs). Conditional attributes vary according to each specific case of decision-making to measure their impact on DMUs efficiencies.

2.3. Step 3: Determine the reducts through Rough Set Theory

The central concept of feature selection technique is called reduct that can be described as a minimum subset of attributes, completely discriminating all objects in a data set. Decision tables with input and output attributes are the best case for reduct application. In this situation, we compute reducts according to the value of the output or decision attributes. Considering the data in Table 1 with four input attributes ($F1$–$F4$) and one output feature with three possible answers, we can obtain a reduct containing the single feature of $F2$ that can cluster decision class into two subsets of objects according to $F2$ value in data set. One of the goals of feature selection technique is to find reducts with the smallest number of attributes. Total number of reducts in an information system with m attributes may be equal to $\binom{m}{[m/2]}$. Computing all reducts is not a simple task and is an Np-hard problem (Banker, Charnes, &

Table 1
Five-object data set.

Object no.	*F1*	*F2*	*F3*	*F4*	*D*
1	0	1	0	2	0
2	1	1	0	2	2
3	0	0	0	1	0
4	0	1	1	0	1
5	0	0	1	3	0

Cooper, 1984). But, fortunately, good heuristic algorithms including genetic algorithm are developed for this reason (Cheng & Titterington, 1994).

2.4. Step 4: Select preferred ANN for each reduct with CVTT

To estimate the quality of constructed neural network we have employed Cross Validation Test Technique (CVTT). In CVTT, the data set is first split into several parts. Then, one part is utilized for testing and the rest are saved for training purpose. These steps are repeated until all parts used as testing set. The final product of CVTT is the mean accuracy of total runs.

As discussed by Cybenko (1989) and Patuwo et al. (1993), a single hidden layer is sufficient in constructing neural nets. Therefore, a single hidden layer neural network is selected in this study. To find the appropriate numbers of hidden nodes in ANN analysis of each reduct, following steps are performed to construct networks with one to q nodes, where q is an optional parameter and will be changed until the desired error is met by the algorithm.[1]

- Training step is performed by scaled conjugate gradient training algorithm Neural Network Toolbox for use with MATLAB1 (2000).
- The model is evaluated by the test data through obtaining MAPE[2] error. There are four basic error estimation methods, which are: Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). All methods, except MAPE have scaled output. Therefore, MAPE method is the most suitable method to estimate the relative error (Azadeh, Ghaderi, Anvari, & Saberi, 2006a, 2006b; Azadeh, Ghaderi, Anvari, & Saberi, 2007; Azadeh, Ghaderi, Anvari, Saberi, & Izadbakhsh, 2007; Azadeh, Ghaderi, & Sohrabkhani, 2007; Azadeh, Ghaderi, Tarverdian, & Saberi, 2006, 2007).

2.4.1. Neural network modelling

Usually train data set contains 70–90% of all data and remaining data are used for test data set (Azadeh et al., 2007). One of the problems that occur during neural network training is called overfitting. The error on the training set is driven to a very small value, but when new data is presented to the network the error is large. The network memorizes the training examples, but it does not learn to generalize for new situations. Early stopping method is used for this problem. In this method, the available data is divided into two subsets. The first subset is the training set, which is used for computing the gradient and updating the network weights and biases. The second subset is the validation set. 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. However, when the network begins to overfit 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. The test set error is not used during the training, but it is used to compare different models. It is also useful to plot the test set error during the training process. If the error in the test set reaches a minimum at a significantly different iteration number than the validation set error, this may indicate a poor division of the data set.

2.5. Step 5: Select the best reducts by ANN results through DEA

Much of the emphasis here is selecting a good subset of conditional attributes according to different data sets constructed by RST. As all reducts have classification quality of 100% in training data set, we may differentiate their performance in the form of accuracy of prediction on unseen data. As selected reducts must be used in predicting system in last step of algorithm, the best reduct is selected when ANN performance is at its best with the selected reduct. This approach can help identify the best subset of parameters effecting DMUs efficiencies.

We have utilized DEA method to rank reducts in this section. As we do not know which attributes are more important for efficiency analysis, it makes sense to use this method. Since mean scores of error ($AERR_i$) alone may not be appropriate in ranking reducts, other variables, including variance ($VarERR_i$), minimum ($MinERR_i$) and maximum ($MaxERR_i$) of error rate for each reduct are considered. DEA method will effectively take into account these variables to select a good subset of attributes after calculating reduct's efficiency score. We treated $AERR_i$, $VarERR_i$, $MaxERR_i$ and $MinERR_i$ for each reduct as inputs of efficiency with the specific output of 1 to calculate efficiency of each reduct. We seek to use ANN result and related information as much as possible with aid of mentioned variables and DEA. MAPE_{ijk} is defined as Mean Absolute Percentage Error of ANN, with regard to kth part of data, used as test data and node j in the hidden layer, both related to ith reduct. Variable ERR_{ij} is defined as $ERR_{ij} = \mathrm{Average}\{\mathrm{MAPE}_{ijk}\colon k = 1, 2, \ldots\}$ or average of MAPE in all constructed ANN for ith reduct with regard to j node in hidden layer. $AERR_i$ is defined as: $AERR_i = \mathrm{Average}(ERR_{ij})$. $VarERR_i$ is defined as: $VarERR_i = Variance(ERR_{ij})$. $MaxERR_i$ is defined as: $MaxERR_i = Max(ERR_{ij})$. $MinERR_i$ is defined as: $MinERR_i = Min(ERR_{ij})$.

2.6. Step 6: Forecasting DMUs efficiencies by selected attributes through ANN

Finally, by selecting the best reduct, the constructed ANN for that reduct with maximum accuracy can be utilized for forecasting DMUs efficiency. This reduction in attributes number decreases the time of decision-making process and consequently reduces the cost of efficiency analysis. By constructed predicting system, managers can assess the new situations before their actual occurrence. They can further manage their human resources through this online decision-making system.

3. The case study

The proposed integrated approach is applied to a large private bank. The case study focuses on 102 branches to analyze the effect of personnel attributes on efficiency of the branches. The proposed algorithm with explanation of each stage is described in the following sections.

3.1. DMU's efficiency calculation with DEA

Athanassopoulos (1997) discuses two models of intermediation and production for financial firms. Intermediation institution collects deposits as input by placing loans as output to make profit. Berger and Humphrey (1991) present examples of intermediation firms. Production model utilizes physical resources as input and collected deposits and loans as outputs (Soteriou & Zenios, 1999). In this paper, we have calculated efficiency on major attributes of production model according to the nature of financial firms in this bank. Table 2 shows DEA inputs, outputs, and Table 3 shows efficiency scores calculated for 102 branches.

[1] In this study the value of the desired minimum error has been defined between 6% and 8% (94–92% confidences) and the value of q has been defined 20. The error is estimated by Mean Absolute Percentage Error (MAPE).

[2] Mean Absolute Percentage Error $\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\mathrm{Actualvalue}_i - \mathrm{Setpointvalue}_i}{\mathrm{Setpointvalue}_i}\right|$ (N: the number of rows).

Table 2
DEA inputs and outputs.

DEA inputs	DEA outputs
Number of employees	Deposits
Fixed assets	Operating income
	Loans

Output-oriented BCC Model is used for efficiency calculation. It is based on maximization of the following objective function:

$$\begin{aligned}
&\text{Max} \quad \theta \\
&\text{s.t.} \quad x_{io} \geqslant \sum_{j=1}^{n} \lambda_j x_{ij}, \quad i = 1, \ldots 2 \\
&\qquad \theta y_{ro} \leqslant \sum_{j=1}^{n} \lambda_j y_{rj}, \quad r = 1, \ldots, 3 \\
&\qquad \sum_{j=1}^{n} \lambda_j = 1 \\
&\qquad \lambda_j \geqslant 0, \quad j = 1, \ldots, 102
\end{aligned} \tag{7}$$

3.2. Decision system definition

We have identified four groups of personnel in each branch. First group are associated as tellers who conduct most of a bank's routine transactions. Among the responsibilities of tellers are accepting deposits, loan payments, and processing withdrawals. They also may sell savings bonds; accept payment for customers' utility bills, charge cards and process necessary paperwork for certificates of deposit. Some tellers specialize in handling foreign currencies or commercial or business accounts. The second group consists of supervisors who cash checks and perform control tasks on transactions. Branch mangers and their assistants are in third and fourth groups. Each branch has one branch manager and may have several personnel assigned to other groups. We have recognized 28 conditional personnel attributes. Fig. 2 shows the relationship between personnel groups and defined personnel specifications.

3.2.1. Personnel attribute categories

Personnel attributes are categorized in four groups as follows:

(1) **Gender:** Female or male.
(2) **Education:** There are five categories of degrees: Below high school diploma, high school diploma, AB, BS, MS and above.
(3) **Age:** This attribute shows the age of the employees.
(4) **Work experience:** Work experience is the experience in a specific field or occupation.
(5) **Others:** Number of singles and tellers.

The decision System is developed by the above attributes. It should be noted that the number of males, singles and tellers are divided by number of personnel in each branch. Also, the number personnel with in each degree category are divided by number of personnel in each branch. Minimum, average and maximum work experience of assistant branch managers, supervisors and tellers are divided by maximum work experience existing in each branch. The data with respect to age do not need to be changed because they are comparable within its category. With regard to the above definitions, the percentage of early-defined attributes may be identified. For example, the percentage of male personnel in one branch is more useful than the number of male personnel in the same branch. In addition, new attributes (percentages) would satisfy the need of ANN to normalize its training data (Cheng & Titterington, 1994).

The above definitions would also show the lack of some attributes in the decision system. For example, percentage of females is not considered in the decision system. Moreover, the availability of the percentage of male personnel in one branch would simplify the percentage of female personnel in the same branch. This is also true for other attributes.

3.3. Determination of the reducts through RST

Twelve reducts were extracted which are shown in Table 4.[3] It can be simply seen that the number's of males has maximum frequency and maximum age of assistant branch managers has minimum frequency in the reduct set.

3.4. Calculate ANN performance for each reduct with CVTT

Preferred ANN is selected with the aid of the error variable. In order to calculate error value, Cross Validation Test Technique (CVTT) is employed with 4-fold. The data set is first split into 4 divisions and then one division is taken as validation and test sample and the remainder becomes the training set. In this study the value of the desired minimum error has been defined between 6% and 8% (92–94% confidence) and the value of q has been defined as 40. The error term is estimated by Mean Absolute Percentage Error (MAPE). Table 5 shows error variables of MAPE_{ijk} calculated for each reduct versus number of nodes in the hidden layer. Also, architect of each preferred ANN is shown in Fig. 3(a-l).[4]

3.5. Select the best reducts by ANN results through DEA

We have utilized DEA method instead to rank reducts in this section. Since mean scores of error ($AERR_i$) alone may not be appropriate in ranking reducts, other variables, involving variance ($VarERR_i$), minimum ($MinERR_{ij}$) and maximum ($MaxERR_{ij}$) of error rate are considered for each reduct. Error variables were shown in Table 5. DEA method will effectively take into account the values of these variables to select a good subset of attributes after calculating reduct's efficiency score. We treated value of $AERR_i$, $VarERR_i$, $MaxERR_i$ and $MinERR_i$ for each reduct as inputs of efficiency with the specific output of 1 to calculate efficiency of each reduct. Calculated full rank efficiency scores along with reduct's ranks are shown in Table 6. As shown in Table 6, the 9th reduct is identified as the best reduct. Number of singles, maximum age of tellers and work experience of branch managers have the greatest impact on efficiency of bank branches. By constructing neural network on 9th reduct with two neurons in hidden layer (Fig. 3(i)), we will obtain an effective expert system, which can be utilized by senior managers of the bank for sensitivity analysis or efficiency prediction of inefficient or new bank branches according to their personnel specifications values.

3.6. Forecasting DMU's efficiencies by selected attributes through ANN

As mentioned, the 9th reduct was selected as the preferred reduct. Number of singles, maximum age of tellers and work experience of branch manager are attributes of 9th reduct. This reduction in number of attributes decreases the time of decision-making process and consequently reduces the cost of efficiency analysis. Seven arbitrary values for present inputs are used as ANN inputs. Table 7 shows the output of ANN.

[3] This method is implemented with RSES Software.
[4] Related Networks are implemented with Neural Network Toolbox of MATLAB.

Table 3
Bank branch's efficiency scores.

Branch ID	BCC eff.	Branch ID	BCC eff.	Branch ID	BCC eff.	Branch ID	BCC eff.	Branch ID	BCC eff.
1	1.00	22	0.96	43	0.93	64	0.87	85	0.94
2	1.00	23	1.00	44	0.9	65	0.87	86	0.89
3	1.00	24	0.9	45	0.81	66	0.91	87	0.88
4	1.00	25	0.92	46	1.00	67	0.92	88	1.00
5	1.00	26	1.00	47	0.70	68	1.00	89	0.91
6	1.00	27	1.00	48	1.00	69	0.93	90	0.84
7	1.00	28	0.93	49	0.71	70	0.91	91	0.91
8	1.00	29	0.89	50	1.00	71	0.89	92	0.82
9	1.00	30	1.00	51	1.00	72	0.87	93	0.82
10	1.00	31	0.95	52	0.95	73	0.93	94	0.80
11	0.98	32	1.00	53	0.72	74	0.93	95	0.94
12	1.00	33	0.92	54	1.00	75	0.93	96	0.74
13	0.94	34	0.96	55	0.89	76	1.00	97	0.81
14	1.00	35	0.95	56	0.94	77	0.89	98	0.72
15	1.00	36	0.86	57	0.91	78	0.93	99	0.82
16	1.00	37	0.98	58	1.00	79	0.95	100	0.69
17	0.88	38	0.99	59	0.96	80	0.82	101	0.88
18	0.94	39	0.89	60	1.00	81	0.99	102	0.71
19	1.00	40	0.97	61	0.94	82	0.92		
20	0.84	41	0.91	62	0.94	83	0.91		
21	0.92	42	1.00	63	1.00	84	0.85		

Eff. = efficiency.

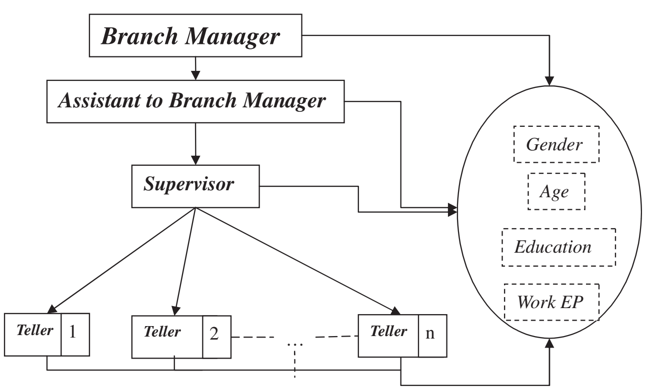

Fig. 2. Relationship between personnel groups and defined personnel specification.

Table 5
Mean Absolute Percentage Error as DEA inputs.

Reduct no.	Average of error (*AERR*)	Variance of error (*VarERR*)	Maximum of error (*MaxERR*)	Minimum of error (*MinERR*)
1	0.2529	0.0339	0.6974	0.1042
2	0.1933	0.0086	0.4147	0.1121
3	0.2014	0.0070	0.3993	0.1045
4	0.1841	0.0105	0.4645	0.0938
5	0.2084	0.0109	0.4872	0.1016
6	0.2634	0.0241	0.6350	0.0995
7	0.2319	0.0206	0.6834	0.1213
8	0.1923	0.0084	0.4693	0.1132
9	0.2009	0.0039	0.3099	0.1291
10	0.1820	0.0077	0.4095	0.1142
11	0.2865	0.0122	0.5380	0.1648
12	0.2013	0.0061	0.3493	0.1179

4. Conclusion and future work

The proposed approach of this paper provided a six-stage analysis to help managers formulate an effective decision-making procedure to demonstrate critical attributes affecting personnel efficiency in particular and total efficiency in general. The purpose is to alert management to the important attributes that should be considered if an effective decision to enhance total efficiency is to be formulated. Determination of critical personnel attributes is a useful procedure to overcome complication associated with multiple inputs and outputs. The proposed algorithm assesses the impact of personnel efficiency attributes on total efficiency through Data Envelopment Analysis (DEA), Artificial Neural Network and Rough Set Theory (RST). The outcome helps managers to construct

Table 4
The set of reducts in the case study.

Reduct ID	Reducts	Reduct size
1	{average work experience of supervisors, average age of tellers, number of males}	3
2	{number of males, average age of supervisors, average age of tellers}	3
3	{number of males, number of tellers, average work experience of assistant managers}	3
4	{minimum age of tellers, maximum work experience of tellers, minimum work experience of assistant managers}	3
5	{number of males, work experience of branch manager, minimum work experience of supervisors}	3
6	{number of males, maximum age of supervisors, maximum work experience of supervisors}	3
7	{number of singles, number of bachelor of science degrees, average work experience of tellers}	3
8	{maximum age of supervisors, average work experience of assistant managers, maximum work experience of supervisors}	3
9	{number of singles, maximum age of tellers, work experience of branch manager}	3
10	{work experience of branch manager, minimum age of supervisors, maximum age of assistant managers}	3
11	{average work experience of tellers, number of bachelor of science degrees, number of singles, number of males}	4
12	{number of tellers, number of bachelor of science degrees, number of singles, number of males}	4

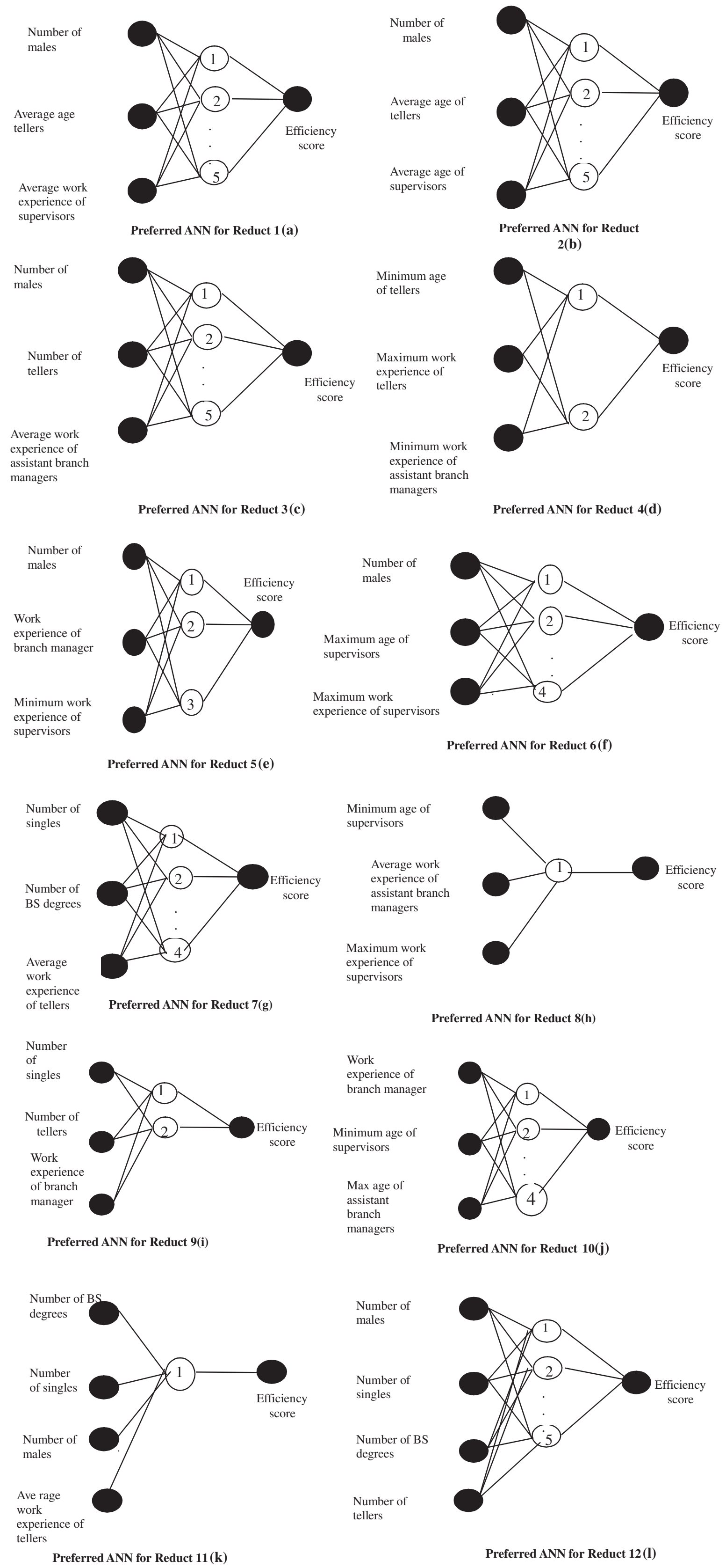

Fig. 3. Architect of the preferred ANN.

helpful system to forecast DMUs efficiencies by selected attributes. Also, this is the first study in literature in which neural networks, Data Envelopment Analysis and Rough Set Theory are integrated for assessment of personnel efficiency. The significant features of

Table 6
DEA scores and ranks.

	Full rank efficiency	Rank
Reduct 9	1.058994	1
Reduct 4	1.037797	2
Reduct 10	1.014341	3
Reduct 3	1.011560	4
Reduct 12	1.000493	5
Reduct 2	0.990344	6
Reduct 8	0.987629	7
Reduct 5	0.979080	8
Reduct 6	0.977991	9
Reduct 1	0.959753	10
Reduct 7	0.898111	11
Reduct 11	0.843254	12

Table 7
Output of ANN with arbitrary input (reduct 9).

	Number of singles	Maximum age of tellers	Work experience of branch manager	Output value (efficiency scores)
1	0.2	0.4	0.7	0.912139
2	0.5	0.7	0.4	0.933977
3	0.4	0.5	0.6	0.925930
4	0.7	0.6	0.7	0.965406
5	0.2	0.3	0.2	0.845798
6	0.2	0.8	0.1	0.894381
7	0.5	0.2	0.4	0.876782

Table 8
The significant features of the proposed algorithm versus existing models and algorithms.

Existing models	Significant features						
	Efficiency computation	Forecasting capability	Decision support system capability	Nonlinear combination	Identification of influentional features	Extraction feature	Ranking capability
The proposed algorithm	√	√	√	√	√	√	√
Data Envelopment Analysis	√						√
Artificial Neural Network	√	√	√				
RST–DEA	√			√		√	
DEA–ANN	√	√	√				
PCA–DEA–ANN	√	√				√	√
FA–DEA–ANN	√	√				√	√

the proposed algorithm in comparison with existing models and algorithms are shown in Table 8. Obviously, the proposed algorithm is superior to the conventional and existing models and algorithms.

The integrated approach was successfully applied to 102 branches of a large private bank. It evaluated the personnel attributes impact total efficiency of bank branches. The results showed that four subsets of conditional attributes with total number of nine attributes from 28 attributes have a critical impact on the accuracy of the optimal solution. This reduction in attributes number decrease the time for decision-making process and consequently reduces the cost of efficiency evaluation.

References

Athanassopoulos, A. D. (1997). Service quality and operating efficiency synergies for management control in the provision of financial services: Evidence from Greek bank branches. *European Journal of Operational Research, 98*, 300–313.

Athanassopoulos, A. D., & Curram, S. P. (1996). A comparison of data envelopment analysis and artificial neural networks as tool for assessing the efficiency of decision making units. *Journal of the Operational Research Society, 47*, 1000–1016.

Azadeh, A., Amalnick, M. S., Ghaderi, S. F., & Asadzadeh, S. M. (2007). An integrated DEA PCA numerical taxonomy approach for energy efficiency assessment and consumption optimization in energy intensive manufacturing sectors. *Energy Policy, 35*, 3792–3806.

Azadeh, A., Anvari, M., & Saberi, M. (2007). An intelligent algorithm for performance assessment of non-homogeneous decision making units. In *Ninth US national congress on computational mechanics, Berkley, USA*.

Azadeh, A., Anvari, M., & Saberi, M. (2008). *An integrated ANN fuzzy C-means normalization algorithm for performance assessment of decision making units*. Technical Report, University of Tehran, Iran.

Azadeh, A., Ghaderi, S. F., Anvari, M., & Saberi, M. (2006a). Measuring performance electric power generations using artificial neural networks and fuzzy clustering. In G. A. Capolino, & L. G. Franquelo (Eds.), *Proceedings of the 32nd annual conference of the IEEE industrial electronics society, IECON, Paris, France*.

Azadeh, A., Ghaderi, S. F., Anvari, M., & Saberi, M. (2006b). Measuring performance electric power generations using artificial neural networks and fuzzy clustering. In *Proceeding of the 32nd annual conference of the IEEE industrial electronics society, IECON, Paris, France*.

Azadeh, A., Ghaderi, S. F., Anvari, M., & Saberi, M. (2007). Performance assessment of electric power generations using an adaptive neural network algorithm. *Energy Policy, 35*, 3155–3166.

Azadeh, A., Ghaderi, S. F., Anvari, M., Saberi, M., & Izadbakhsh, H. (2007). An integrated artificial neural network and fuzzy clustering algorithm for

performance assessment of decision making units. *Applied Mathematics and Computation, 187*, 584–599.
Azadeh, A., Ghaderi, S. F., & Izadbakhsh, H. (2007). Integration of DEA and AHP with computer simulation for railway system improvement and optimization. *Applied Mathematics and Computation, 195*, 775–785.
Azadeh, A., Ghaderi, S. F., & Sohrabkhani, S. (2007). Forecasting electrical consumption by integration of neural network, time series and ANOVA. *Applied Mathematics and Computation, 186*, 1753–1761.
Azadeh, A., Ghaderi, S. F., Tarverdian, S., & Saberi, M. (2006). Integration of artificial neural networks and GA to predict electrical energy consumption. In G. A Capolino, & L. G. Franquelo (Eds.), *Proceedings of the 32nd annual conference of the IEEE industrial electronics society – IECON'06, Conservatoire National des Arts and Metiers, Paris, France*.
Azadeh, A., Ghaderi, S. F., Tarverdian, S., & Saberi, M. (2007). Integration of artificial neural networks and genetic algorithm to predict electrical energy consumption. *Applied Mathematics and Computation, 186*, 1731–1741.
Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in DEA. *Management Science, 32*, 1078–1092.
Bauer, P. W. (1990). Recent developments in the econometric estimation of frontiers. *Journal of Econometrics, 46*, 39–56.
Becerra-Fernandez, I., Zanakis, S. H., & Walczak, S. (2002). Knowledge discovery techniques for predicting country investment risk. *Computers and Industrial Engineering, 43*, 787–800.
Berger, A. N., & Humphrey, D. B. (1991). The dominance of inefficiencies over scale and product mix economies in banking. *Journal of Monetary Economics, 28*, 117–148.
Beynon, M. J., & Peel, M. J. (2001). Variable precision rough set theory and data discretion: An application to corporate failure prediction. *Omega, 29*, 561–576.
Błaszczyński, J., Greco, S., & Słowiński, R. (2007). Multi-criteria classification – A new scheme for application of dominance-based decision rules. *European Journal of Operational Research, 181*, 1030–1044.
Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of DMUs. *European Journal of Operational Research, 2*, 429–444.
Charnes, A., Cooper, W. W., Lewin, A. Y., & Seiford, L. M. (1994). *Data Envelopment Analysis: Theory, Methodology and Applications*. Boston, MA: Kluwer Academic Publishers.
Cheng, B., & Titterington, D. M. (1994). Neural networks: A review from a statistical perspective. *Statistical Science, 9*, 2–54.
Chou, H. C., Cheng, C. H., & Chang, J. R. (2007). Extracting drug utilization knowledge using self-organizing map and rough set theory. *Expert Systems with Applications, 33*, 499–508.
Costa, A., & Markellos, R. N. (1997). Evaluating public transport efficiency with neural network models. *Transportation Research C, 5*, 301–312.
Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems, 2*, 303–314.
Fan, T. F., Liu, D. R., & Tzeng, G. H. (2007). Rough set-based logics for multi-criteria decision analysis. *European Journal of Operational Research, 182*, 340–355.
Hassanien, A. (2007). Fuzzy rough sets hybrid scheme for breast cancer detection. *Image and Vision Computing, 25*, 172 183.
Inuiguchi, M., & Miyajima, T. (2007). Rough set based rule induction from two decision tables. *European Journal of Operational Research, 181*, 1540–1553.
Kowalczyk, W., & Slisser, F. (1997). Modeling customer retention with rough data models. In *Proceedings of the 1st European symposium PKDD '97, Trondheim, Norway* (pp. 4–13).
Kusiak, A., & Tseng, B. T. L. (2000). Data mining in engineering design: A case study. In *IEEE international conference on robotics and automation, proceedings* (Vol. 1, pp. 206 – 211).
Kusiak, A., Kern, J. A., Kernstine, K. H., & Tseng, B. T. L. (2000). Autonomous decision-making: A data mining approach. *IEEE Transactions on Information Technology in Biomedicine, 4*, 274–284.
Lili, W., & Zhi, P. (2001). Fuzziness in rough sets based on general relations. *Journal of Ningxia University (Natural Science), 22*, 352–355.
Liu, M., Chen, D., Wu, C., & Li, H. (2006). Reduction method based on a new fuzzy rough set in fuzzy information system and its applications to scheduling problems. *Computers and Mathematics with Applications, 51*, 1571–1584.
Neural Network Toolbox for use with MATLAB1 (2000). The Math-Works Inc., USA.
Patuwo, E., Hu, M. Y., & Hung, M. S. (1993). Two-group classification using neural networks. *Decision Sciences, 24*, 825–845.
Pawlak, Z. (1982). Rough sets. *International Journal of Computer and Information Sciences, 11*, 314–356.
Pawlak, Z. (1991). *Rough sets: Theoretical aspects of reasoning about data*. Boston, MA: Kluwer.
Pendharkar, P. C., & Rodger, J. A. (2003). Technical efficiency-based selection of learning cases to improve forecasting accuracy of neural networks under monotonicity assumption. *Decision Support Systems, 36*, 117–136.
Ravi Kumar, P., & Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques – A review. *European Journal of Operational Research, 180*, 1–28.
Ruhe, G. (1996). Qualitative analysis of software engineering data using rough sets. In *Proceedings of the 4th international workshop on rough sets, fuzzy sets, and machine discovery, Tokyo, Japan* (pp. 292–299).
Sanchis, A., Segovia, M. J., Gil, J. A., Heras, A., & Vilar, J. L. (2007). Rough Sets and the role of the monetary policy in financial stability (macroeconomic problem) and the prediction of insolvency in insurance sector (microeconomic problem). *European Journal of Operational Research, 181*, 1554–1573.
Santin, D., & Valino, A. (2000). Artificial neural networks for measuring technical efficiency in schools with a two-level model: An alternative approach. In *II Oviedo workshop on efficiency and productivity, Oviedo*.
Santin, D., Delgado, F. J., & Valino, A. (2004). The measurement of technical efficiency: A neural network approach. *Applied Economics, 36*, 627–635.
Serrano Cinca, C., Mar Molinero, C., & Chaparro García, F. (2002). Behind DEA efficiency in financial institutions. *Accounting and Finance, 40*, 30–45.
Shyng, J. Y., Wang, F. K., Tzeng, G. H., & Wu, K. S. (2007). Rough Set Theory in analyzing the attributes of combination values for the insurance market. *Expert Systems with Applications, 32*, 56–64.
Sikder, I. U., & Gangopadhyay, A. (2007). Managing uncertainty in location services using rough set and evidence theory. *Expert Systems with Applications, 32*, 386–396.
Soteriou, A., & Zenios, S. A. (1999). Operations, quality and profitability in the provision of banking services. *Management Science, 45*, 1221–1238.
Stefanowski, J., & Slowinski, K. (1997). Rough sets as a tool for studying attribute dependencies in the urinary stones treatment data set. In *Rough sets and data mining: analysis and imprecise data* (pp. 177–196).
Triantaphyllou, E., Liao, T. W., & Iyengar, S. S. (2002). A focused issue on data mining and knowledge discovery in industrial engineering. *Computers and Industrial Engineering, 43*, 657–659.
Tseng, T. L. B., & Huang, C. H. (2007). Rough set-based approach to feature selection in customer relationship management. *Omega, 35*, 365–383.
Tsumoto, S. (1997). Extraction of experts' decision process from clinical databases using rough set model. In *Proceedings of the 1st European Symposium PKDD '97, Trondheim, Norway* (pp. 58–67).
Wang, X., Yang, J., Jensen, R., & Liu, X. (2006). Rough set feature selection and rule induction for prediction of malignancy degree in brain glioma. *Computer Methods and Programs in Biomedicine, 83*, 147–156.
Wu, D., Yang, Z., & Liang, L. (2006). Using DEA-neural network approach to evaluate branch efficiency of a large Canadian bank. *Expert Systems with Applications, 31*, 108–115.
Xia, W., & Wu, Z. (2007). Supplier selection with multiple criteria in volume discount environments. *Omega, 35*, 494–504.
Zhai, L. Y., Khoo, L. P., & Fok, S. C. (2002). Feature extraction using rough set theory and genetic algorithms—An application for the simplification of product quality evaluation. *Computers and Industrial Engineering, 43*, 661–676.
Ziarko, W. (1993). Variable precision rough set model. *Journal of Computer and System Sciences, 46*, 39–59.