Author's Accepted Manuscript

www.elsevier.com/locate/omega

DEANN: A Healthcare Analytic Methodology of Data Envelopment Analysis and Artificial Neural Networks for the Prediction of Organ Recipient Functional Status

Nicholas Misiunas, Asil Oztekin, Yao Chen, Kavitha Chandra

PII: S0305-0483(15)00062-6
DOI: http://dx.doi.org/10.1016/j.omega.2015.03.010
Reference: OME1513

To appear in: *Omega*

Received date: 24 October 2014
Accepted date: 28 March 2015

Cite this article as: Nicholas Misiunas, Asil Oztekin, Yao Chen, Kavitha Chandra, DEANN: A Healthcare Analytic Methodology of Data Envelopment Analysis and Artificial Neural Networks for the Prediction of Organ Recipient Functional Status, *Omega*, http://dx.doi.org/10.1016/j.omega.2015.03.010

DEANN: A Healthcare Analytic Methodology of Data Envelopment Analysis and Artificial Neural Networks for the Prediction of Organ Recipient Functional Status

Nicholas Misiunas[a], Asil Oztekin[b], Yao Chen[b], Kavitha Chandra[a]

[a]*Department of Electrical and Computer Engineering, Francis College of Engineering, University of Massachusetts Lowell, MA, 01854 USA.*

[b]*Department of Operations and Information Systems, Manning School of Business, University of Massachusetts Lowell, MA, 01854 USA.*

Abstract

The problem of effectively preprocessing a dataset containing a large number of performance metrics and an even larger number of records is crucial when utilizing an ANN. As such, this study proposes deploying DEA to preprocess the data to remove outliers and hence, preserve monotonicity as well as to reduce the size of the dataset used to train the ANN. The results of this novel data analytic approach, i.e. DEANN, proved that the accuracy of the ANN can be maintained while the size of the training dataset is significantly reduced. DEANN methodology is implemented via the problem of predicting the functional status of patients in organ transplant operations. The results yielded are very promising which validates the proposed method.

Keywords: data envelopment analysis (DEA), artificial neural networks (ANN), training dataset reduction, stratification of efficiency layers, healthcare analytics, organ transplant

1. Introduction

Organ transplants are one of the most viable treatment options for patients with organ failures and may also be their only option. Coupled with the cost of the operation and the lack of readily available organs, the need for pairings of organ donors and recipients which result in successful transplants is critical. However, datasets on donors and recipients may contain a vast amount of information, both in number of records and in attributes of the donor, recipient, and their relationship. The need to parse this data is therefore, acute and any single person attempting to perform a prediction may result in heavy bias as the consideration of the important attributes may be difficult, both due to their number and the way in which it may be difficult to determine which attributes contribute towards the outcome of the transplant. There exists a need, therefore, to not only perform accurate predictions on a complex dataset, but also to parse this dataset in some way so as to reduce it to a manageable form. Investigations into this have been performed using different approaches such as by Oztekin *et al*. (2011), who analyzed a lung transplant dataset using decision trees, and reinforced by studies as that by Zhuang *et al.*(2009) who demonstrate the effectiveness of data mining and machine learning for decision making by medical practitioners. Meisel and Mattfeld (2010) also supported this idea by identifying key areas in which operational research and data mining can work synergistically to create innovative approaches towards solutions for problems concerning decision making.

As the data here is ill understood, however, a prediction method that can cope with this lack of knowledge must be utilized. Artificial neural networks (ANN) are one tool that are capable of being trained on a dataset containing attributes of which relationships may be rather complex and are capable of performing accurate predictions on a testing set. ANNs might suffer from over-fitting, however, and may be sensitive towards data that contains conflicting observations, and therefore the dataset would be preprocessed by data envelopment analysis (DEA), a linear programming method for determining the relative efficiency of a set of observations. This would hypothetically allow a reduction in the dataset used for training the ANN without greatly impacting its performance.

ANNs consist of layers of neurons with distinct weights separating the neuron connections which allow the ANN to be trained on a complex dataset and determine its own understanding of the relationship of the attributes. The medical field is one particularly well-suited application of ANNs, for example the predictions of organ transplants by Dvorchik *et al*. (1996), cancer diagnosis by Abass (2002), or other clinical applications as shown by Dybowski and Gant (2001).

On the other hand, DEA is a linear programming method for evaluating the relative efficiency of a set of decision making units (DMU) by analyzing their weighted inputs and outputs. This method is flexible and allows the performance/efficiency of these DMUs to be analyzed based upon a set of selected performance metrics. Seiford and Zhu (2003) show how DMUs can be stratified, separating them into layers of efficiency, which would be useful for a complex dataset (such as the one analyzed in this study). In this work DMUs are individual transplant records. Pre- and post-transplant variables may be considered as inputs and outputs, respectively. Inefficient records may then be removed from the dataset as a preprocessing method. Kotsiantis (2006) discusses the importance of data preprocessing for machine learning and states that it is this area in which DEA would hypothetically contribute and Liu *et al* (2013) demonstrate the high growth rate of DEA literature.

These two methods would allow the poorly understood transplant dataset to be preprocessed so that the prediction method, in this case the ANN, is trained only on the most efficient data, which renders the prediction to be more efficient and effective. This would prevent contradictive data from being fed onto the ANN and weakening its understanding of the relationship amongst the performance metrics.

The remainder of this paper is organized as follows: Section 2 outlines a background on DEA and ANN as well as covering the proposed hybrid methodology of this study and how they are integrated with respect to the dataset. Section 3 presents the results of the study performed and the strengths and weaknesses of this data analytic approach. Section 4 concludes the paper and also briefly mentions critical future research considerations.

2. Literature Review

Several research studies have been performed in different areas both from a DEA and an ANN perspective. A thorough comparison of the benefits and drawbacks of both DEA and ANNs was performed by Athanassopoulos and Curram (1996). They showed that one of the prime advantages the ANN has over strict DEA is that there is no causality required, which is very important in an organ transplant setting as in the current study. ANNs also allow a combination of continuous and discrete data to be considered without further modifications that must be done for DEA. These advantages make ANN a suitable component in a two-step hybrid methodology composed of DEA and ANNs, as the benefits of DEA can be utilized to process the data and the ANN can perform the predictions, which draws upon the advantages of both. A hybrid approach was also performed by Samoilenko and Osei-Bryson (2010) in which DEA and ANNs are utilized in a joint effort to determine the complexity in determining relative efficiency when there are heterogeneous levels of input and output relationships amongst the DMUs.

A non-hybrid approach was performed by Liu *et al*. (2009) who used a super-efficiency DEA model to rank the observations. Super-efficiency DEA model does not result in efficiencies that are constrained to being less than one. This is performed by removing the i^{th} observation from the set of inputs and outputs when the i^{th} set itself is analyzed. This is in contrast to normal DEA operations where every observation is included in each analysis, even the i^{th}. These values are then the goal of the ANN. As such, when the ANN is trained using the set of training observations, these values must lie within a specified distance of the super-efficiency values of the DEA model, otherwise training continues. ANN is, therefore, being trained to perform the operation of the super-efficiency DEA model, and does not fit well suited for situations in which the relationships between the inputs and outputs is ill understood, the very basis in which ANNs should thrive. Furthermore, the super-efficiency DEA model is already an efficient algorithm that can handle large numbers of observations but is, however, fairly restrained when it comes to the missing data, which is the ANNs would be able to handle. This is, however, not investigated here. There is also no mention of preprocessing of the dataset, which may be vital in other applications where there is missing data or large numbers of inputs and outputs. Chen (2010) performed similar work, where pre- and post-predictions are classified using DEA to judge the effectiveness of the ANN. Ozdemir and Temur (2009) trained an ANN to predict the efficiency of the input-oriented DEA.

Pendharkar and Roger (2003) and Pendharkar (2005) utilized DEA as a tool for preprocessing of a dataset that would be used for forecasting in an ANN. The preprocessing is performed in that case in an attempt to enforce monotonicity upon the inputs to the ANN. Doing so, DEA would allow those observations that do not meet this property to be identified and removed, allowing more accurate predictions. More accurate predictions arising from preservation of monotonicity of the input observations was shown by Wang (1995, 2003). More specifically, monotonicity results in a reduction of over-fitting, a perennial problem of prediction models. The DEA approach used in that study takes only those observations that lie on a specific frontier and that satisfy monotonicity. One variant of the DEA model is stratification (Seiford and Zhu, 2003), which allows all observations to be placed on a set of frontiers, where the first frontier contains the most efficient observations; the second contains all those which are the second best, and so on. A single frontier or combination of frontiers may be used to strictly or weakly satisfy monotonicity, respectively. The greater the number of frontiers used (and thus a greater number of inefficient observations), the greater the chance of the ANN developing a non-monotonic prediction. Another hybrid approach was performed by Samoilenko and Osei-Bryson (2013) which utilized DEA to obtain the efficiency of the observations and utilized this along with a neural network and other classification methods to create a decision support system for assessing the performance of organizations.

Olanrewaju *et al*. (2011, 2012) assessed energy efficiency of an industrial sector using an ANN. DEA was then utilized to rank the predictions of the ANN, using as inputs the actual energy consumption and the predictions of the ANN as outputs. This allows the efficiency of the predictions to be seen. DEA only has the knowledge of the predicted and observed energy consumption and thus, the only conclusions that can be made concern those of the accuracy of the ANN. This can, however, allow outliers to be seen in the predictions of the ANN, which is not investigated in that study. This may warrant further investigation, both in the causes of those outliers and the benefits of this ranking in terms of refining accuracy of the ANN.

One variation of DEA is referred to as super-efficiency, which arises when a specific DMU under evaluation is removed from the set being considered. This allows discrimination among the efficient DMUs as their efficiency score would no longer be a maximum of one. Chen (2005) analyzed super-efficiency when infeasibility may arise. Super-efficiency itself is a measure of a DMU's stability in terms of its efficiency and therefore, infeasibility may arise when DMUs are stable to their input/output changes. A model was developed that solves these infeasibility problems, which may be useful for medical data in which a particular organ transplant is invariant towards a specific input or output but still needs to be considered in the wider scope of prediction. Super-efficiency was also studied by Zhu (2001) as a means of judging the sensitivity of a DMU. Super-efficiency was utilized by Nahra *et al*. (2009) for a two-stage analysis of efficiency on data from the National Drug Abuse Treatment Survey.

Another multi-stage process was performed by Kyo *et al*. (2010), where an ANN was not only used to perform a transformation on the dataset to remove the outnumbered inputs and outputs in relation to the number of observations, but also utilized an analytic network process (ANP) in conjunction with an ANN to constrain the weight range in DEA resulted in a more reasonable result, yielding weights which were in a range that were deemed acceptable given the dataset utilized. Although interesting and useful for DEA, in the organ transplant situation the number of observations far outnumbers the number of inputs and outputs and therefore this type of preprocessing is not necessary.

One of the difficulties in using DEA is its difficulty to, or even lack of an ability to, handle fuzzy data or data that is not continuous. A review by Hatami-Marbini *et al*. (2011) reviews fuzzy DEA methods and presents classification schemes considering many works published over a twenty year period. These types of approaches with DEA would be mandatory for the future of a hybrid DEA approach considering such a varied dataset as the organ dataset used here and are of the utmost importance for future research in this area. Cook *et al* (2014) also discuss the ability of DEA to handle mixed and raw data as well as the input and output selection. Matin and Kuosmanen (2009) have developed a foundation for DEA that handles integer valued inputs and outputs, essential for the application in datasets that contain ordinal variables.

Another important parameter to consider in ANNs is the training time. Datasets are not only getting larger but there is an increasing trend towards deep learning and other complex neural networks that provide certain benefits over classical networks, as discussed by Arel and Rose (2010). The need for reduction in dataset size is therefore essential in order to reduce training times. The fact brings a critical issue along with it: the focus on optimal choice of observations in order to obtain high ANN accuracy. The implementation of DEA would hypothetically assist towards this goal.

3. Proposed DEANN Methodology

As the dataset used in this study is rather complex and vast, there are many steps that are involved in this work. Figure 1 illustrates the generic DEANN methodology which can conveniently be used dealing with such huge datasets in various settings. In this study, the data is preprocessed using DEA so that efficient organ transplants may be separated and the ANN trained on these. The ANN is tested on the full dataset to determine its accuracy. The ANN can be retrained if the dataset is updated with new efficient transplants and continuously provide necessary predictions.

The DEANN methodology integrates a data analytic method, i.e. DEA, to parse the large and complex dataset. This processed data is then utilized by an ANN for training and testing purposes. Finally, the impact of various stratified efficiency frontiers on the accuracy and training time of the ANN are analyzed.

3.1 Data Preprocessing Using DEA

In this part of the process the dataset is preprocessed. Selection of variables and deletion of redundant or empty observations are performed before the consideration by DEA. The selection of variables is conducted in light of the literature (Oztekin *et al.*, 2011). Once this step is performed, the result is then fed into DEA and separated into unique stratified efficiency frontiers, where the first frontier represents all observations containing the highest efficiencies (and are all equal). The second frontier represents the second set of efficiencies, and so on. The output is a selection of observations which have been ranked based on their relative efficiencies, where the output is the graft survival time and all other variables are considered as inputs.

3.2 Training and Testing of the ANN

The next part involves training and testing the ANN using the results of DEA. Although DEA, in the preliminary phase, only considers continuous data, the ANN is capable of considering all variables, regardless of their type or complexity, and as such the full list of variables is utilized for training the ANN. The observations are unsorted and ten-fold cross-validation is used for testing of the ANN.

3.3 Analysis of Stratified Efficiency Frontiers on ANN Accuracy

The main objective of the DEANN methodology is to gain an understanding of the effects of preprocessing of DEA on the accuracy of the ANN and this part of the process is the analysis of the effects of the stratified efficiency frontiers on the accuracy of the ANN. The impact of single and multiple layers are investigated. The DEANN method will be validated through the accuracy of the ANN and reduction of training dataset size.

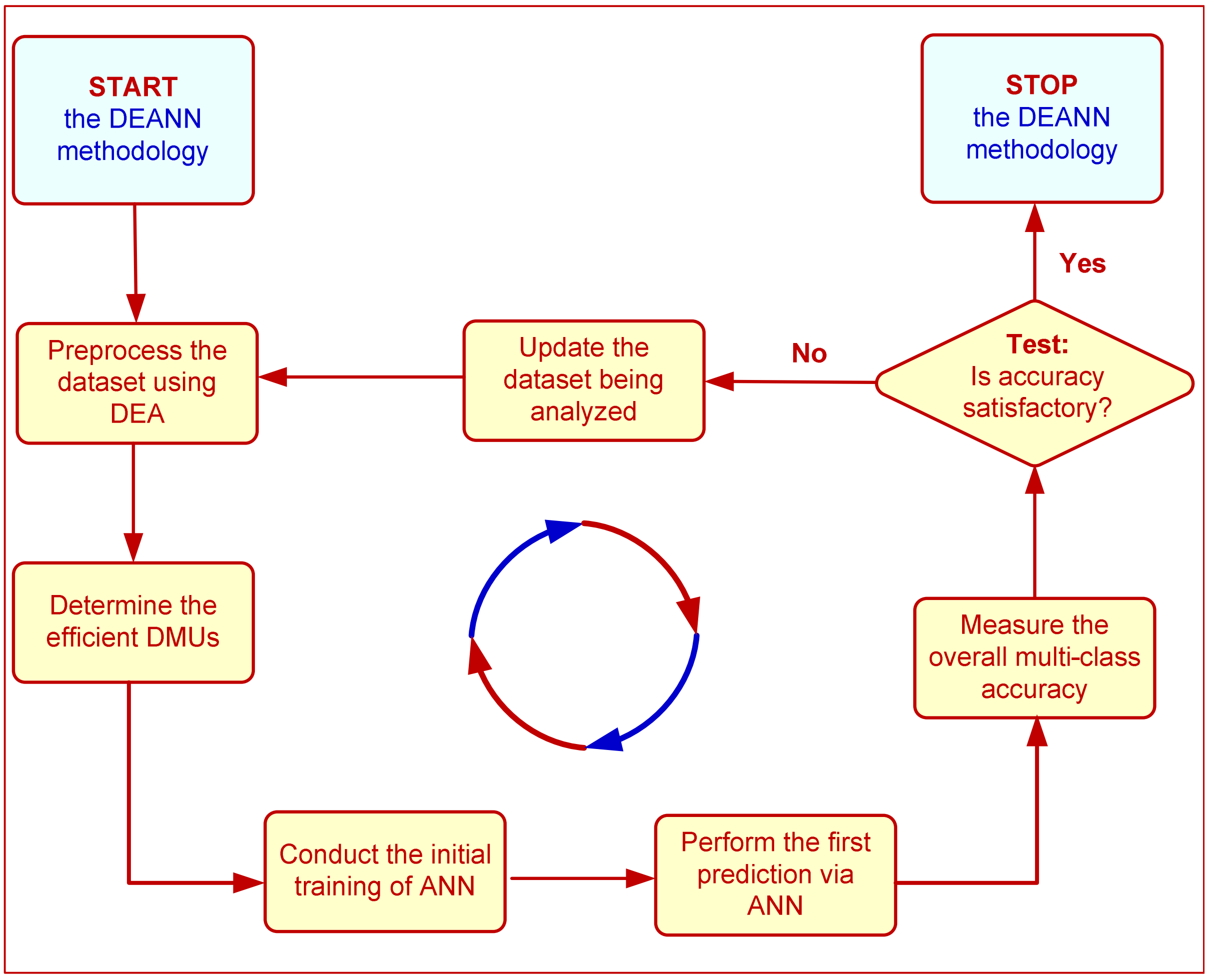

Fig. 1. A pictorial representation of the DEANN methodology

3.4 Overview of Data Preparation

This work examines the Thoracic dataset provided by the United Network for Organ Sharing (UNOS). This dataset consists of 16,771 records and 442 variables containing information on all lung and heart transplants performed in the US that have been reported to

UNOS since October 1, 1987. Each record contains copious amounts of information on both the recipient and donor as well as other metrics such as distance from donor to hospital, date of graft failure, etc. The choice of variables for each segment of this hybrid study as well as the analysis of variables containing "codes" are both discussed next. This work only considers lung transplants in order to simplify predictions as this is the first study in literature utilizing ANN and DEA integration on healthcare data. Any records containing missing data are removed which results in 12,744 records for further processing. DEA reduces this even further but as the ANN is always trained on a subset of the entire dataset that is acceptable for this study.

The choice of variables from the dataset is critical. Oztekin *et al*. (2011) consider their past research (Oztekin *et al.*, 2009; Delen *et al.,* 2010) and select twenty-five variables to consider, aiming to predict the graft survival time and patient functional status. Table 1 lists those variables which are utilized by DEA with inputs in the first section and outputs in the second. DEA is best suited for continuous data and hence a selection of variables which are continuous was performed, though other data types may still be considered by DEA as depicted by Cook and Zhu (2006).

Table 1. Variables Considered by DEA

Attribute	**Description**
*BMI_TRR	BMI of recipient at transplant
*BMI_DON	BMI of donor
AGE	Age of recipient (years)
AGE_DON	Age of donor (years)
DAYSWAIT_CHRON	Days for recipient on waiting list
FUNC_STAT_TRR	Functional status of recipient at transplant
GTIME	Survival time of graft (days)
FUNC_STAT_TRF	Functional status of recipient at follow-up

**Those two attributes were not present in the UNOS dataset, but were calculated in this study.*

It should be noted that both BMI variables were calculated from the height and weight provided in the dataset. The U.S. Renal Data System Coordinating Center (2013) researcher's guide was utilized here. Although the dataset was not obtained through the USRDS, their guide provided important descriptions of variables that were only vaguely stated by UNOS, as well as classifications of codes for certain variables such as FUNC_STAT_TRF, where the values take on numbers which are codes for certain patient statuses. This variable in particular had to be analyzed for it took on not only values of 1,2,3 representing activity limitations of the patient, but also values of 4010, 4020, … , 4100 for functional status of 10%, 20%, …, 100% for adults. The 2000 codes were not utilized here as young adults constituted a small group within the dataset and therefore were excluded. Using these codes, this variable is now represented as 10, 20, …, 100 and can be utilized in the DEA study as a continuous variable. As DEA treats larger values as large “amounts”, all variables excluding GTIME and both FUNC_STAT variables were inverted so that DEA would appropriately consider larger values as better. If this was not performed, DEA would act as though a larger AGE, for instance, should result in a large GTIME, which is in fact counter-intuitive.

Table 2. Variables considered by ANN

Attribute	**Description**
BMI_TRR	BMI of recipient at transplant (calculated)
BMI_DON	BMI of donor (calculated)
AGE	Age of recipient (years)
AGE_DON	Age of donor (years)
DAYSWAIT_CHRON	Days for recipient on waiting list
FUNC_STAT_TRR	Functional status of recipient at transplant
AMAT	A locus match level
BMAT	B locus match level
DRMAT	DR locus match level
HLAMAT	HLA match level
HIST_ALCOHOL_OLD_DON	Deceased donor –history of alcohol dependency
HIST_COCAINE_DON	Deceased donor-history of cocaine use
HIST_HYPERTENS_DON	Deceased donor-history of hypertension
HIST_IV_DRUG_OLD_DON	Deceased donor-history of IV drug use
HIST_CIG_DON	Deceased donor-history of cigarettes
HIST_CANCER_DON	Deceased donor-history of cancer
HIST_DIABETES	Deceased donor-history of diabetes

HIST_MI	Deceased donor-history of Myocardial Infarction
MED_COND_TRR	Recipient medical condition at transplant
ETHCAT_DON	Donor ethnicity category
ETHCAT	Recipient ethnicity category
GENDER_DON	Donor gender
GENDER	Recipient gender
GTIME	Survival time of graft (days)
FUNC_STAT_TRF	Functional status of recipient at follow-up

Table 2 lists the variables that were chosen for the ANN, closely following those chosen by Oztekin *et al*. (2011) with minor differences. No variables are inverted for the ANN as it handles all data types naturally. Ordinal data types are decoded into a set of binary variables whose size is dependent on how many different values the ordinal variable consists of. Continuous variables are normalized to zero mean, unit variance. These result in seventy input variables to the ANN.

The TRR and TRF abbreviations on some variables are used to specify the type of data within the medical database, as certain descriptors may be applied to both recipient and donor, as well as in different time periods. TRR represents a recipient's feature at the time of transplant and TRF represents a feature at a follow-up (after the transplant has occurred). There are many others within the overall database, but they are not discussed as they are not used here.

3.5 Data Analytic Models Deployment

DEANN methodology is composed of two famous data analytic methods, namely Data Envelopment Analysis (DEA) and Artificial Neural Networks (ANN), both of which are briefly outlined in the following subsections.

3.5.1 Data Envelopment Analysis

DEA is a tool that provides rankings of decision making units (DMUs) in terms of relative efficiency given their inputs and outputs (Zhu, 2014). DEA calculates the best frontier and all DMUs lying on the frontier have an efficiency of 1, whilst all others would have an efficiency of less than 1. In a constant returns-to-scale (CRS) situation this method takes the i^{th} DMU and seeks to contract the input vector x_i to the inner-boundary that is in the frontier of the total set of DMUs. The constraints ensure that this point does not lie outside this frontier. θ therefore is an efficiency metric that as stated previously ranges from 0 to 1.

Eq. 1 represents the input-oriented multiplier form for the CRS frontier. K is the number of inputs, M is the number of outputs, and N is the number of DMUs. It can be seen that it aims to maximize the weighted outputs constrained to the weighted inputs and outputs. In this form there are N constraints. For situations in which the number of DMUs is very large (as is the case in medical data in which the number of records vastly outnumbers the total combined inputs and outputs), a transformation of this equation to a form that reduces the constraints to being in terms of the inputs and outputs would be beneficial.

$$\begin{aligned} &\max && \vec{y}_i \vec{\mu} \\ &\text{subject to} && -\boldsymbol{X}\vec{v} + \boldsymbol{Y}\vec{\mu} \le 0 \\ &&& \vec{x}_i \vec{v} = 1 \end{aligned} \tag{1}$$

1. K is the number of inputs
2. M is the number of outputs
3. N is the number of DMUs
4. $\vec{y}_i$ is a $1xM$ vector of outputs for the i^{th} DMU
5. $\vec{x}_i$ is a $1xK$ vector of inputs for the i^{th} DMU
6. $\vec{\mu}$ is a $Mx1$ vector of output weights
7. $\vec{v}$ is a $Kx1$ vector of input weights
8. $\boldsymbol{X}$ is a NxK matrix of inputs
9. $\boldsymbol{Y}$ is a NxM matrix of outputs

Eq. 2 represents the envelopment form of the input-oriented CRS frontier. In this form the raw efficiency metric θ is optimized and there are $K+M$ constraints. This form can also be used to calculate a variable returns-to-scale (VRS) frontier, merely by the addition of a constraint that sums $\vec{\lambda}$ to 1. Slacks can easily be introduced into the model which allow a determination of the amount of slack a specific DMU has for each input. Utilizing both the CRS and VRS efficiencies; the status of the frontier, such as decreasing or increasing returns, can be calculated.

$$\min \quad \theta$$

$$\text{subject to} \quad -\vec{y}_i + \boldsymbol{Y}^T \vec{\lambda} \geq 0 \tag{2}$$

$$\theta \vec{x}_i - \boldsymbol{X}^T \vec{\lambda} \geq 0$$

$$\vec{\lambda} \geq 0$$

1. θ is a scalar

2. $\vec{\lambda}$ is a *Nx1* vector of constants

One variation of DEA allows the DMUs to be classified on distinct efficiency levels. This variation, called stratification, is incredibly useful for preprocessing in which the DMUs are being used as a training dataset for an ANN as it allows different levels to be chosen so that the ANN may be trained with specific levels of efficiency, reducing contradictive training data and improving the accuracy of the ANN.

Eq. 3 is the stratified CRS input-oriented model. Here the efficiency is calculated and all those DMUs which possess an efficiency of 1 (thus lying on the frontier) are removed from the pool and the minimization is run again, repeating until there are no DMUs left. This allows l frontiers to be calculated (Seiford and Zhu, 2003).

$$\min_{\lambda_{j,\theta(l,k)}} \quad \theta(l,k)$$

$$\text{subject to} \quad -\vec{y}_{ji} + \boldsymbol{Y}_j^T \vec{\lambda}_j \geq 0 \tag{3}$$

$$\theta \vec{x}_{ij} - \boldsymbol{X}_j^T \vec{\lambda}_j \geq 0$$

$$\vec{\lambda}_j \geq 0$$

$$j \in F(\boldsymbol{J}^l)$$

The benefit of performing stratification is that it allows a specific layer, or sets of layers, to be chosen and thus those DMUs can be separated from the entire pool. In this way, the ANN may be trained on only the most efficient layer, the two most efficient layers, and so on and so forth, to investigate the accuracy gained by the ANN. These layers allow the ANN to be trained on non-contradictive data containing no outliers, which would improve its accuracy.

3.5.2 Artificial Neural Networks

Artificial Neural Networks provide significant contributions towards the predictions of medical outcomes (Dayhoff & DeLeo, 2001). ANNs themselves are inspired by the networks of living neurons and how they are created. Due to the way in which ANNs are trained through supervised learning, they are highly attractive towards datasets in which there are many complex attributes whose relationships may be ill understood. The medical field is one application of this, such as cancer classification or organ transplant survival prediction, and ANNs have seen a recent resurgence due to this application. An ANN itself is comprised of an input layer, a set of one or more hidden layers, and an output layer, as shown in Figure 2.

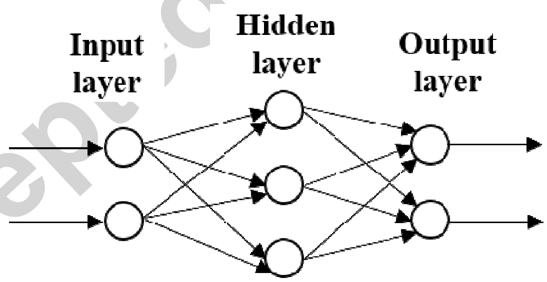

Fig. 2. Representation of a three layer Multi-Layer Perceptron

Each layer has weighted inputs and outputs. Consider a single neuron *j*:

$$Y_j = \sum_i w_{ij} x_i + \theta_j \tag{4}$$

where w_{ij} are the weights assigned to the inputs to that neuron x_i (which are the outputs of the neurons of the previous layer) and θ_j is the bias. The weighted sum Y_j as in Eq. (4) is then passed through a normalization function, such as the Sigmoid function as represented via Eq. (5), which transforms it to be in the range of [0,1]:

$$f\left(x\right)=\frac{1}{1+e^{-x}} \tag{5}$$

The determination of these weights is performed during the training phase, where the network is exposed to training data and adjusts the weights to optimize the classification or prediction performed. The training data must be used in conjunction with a set of testing data, which is an integral part of the training process. If a neural network is trained heavily on a set of data, it may be over-trained and thus, lose its ability to generalize its predictions or classifications towards problems in general.

The training process itself is most commonly performed following a back-propagation algorithms, which is a gradient descent algorithm that propagates the errors through the network. This minimizes the total error by adjusting the weights and the error most conventionally utilized being the root mean square error (RMSE) as in Eq. (6).

$$\mathrm{RMSE}=\sqrt{\frac{1}{n}\sum_{t=1}^{n}(Y_t-O_t)^2} \tag{6}$$

where O_t is the output of the t^{th} unit. Back-propagation training adjusts the weights according to Eq. (7):

$$\Delta w_{ij}=-\eta\frac{\partial\left(\mathrm{RMSE}\right)}{\partial w_{ij}} \tag{7}$$

where η is the parameter governing the magnitude of steps to be performed.

The final task is to determine its accuracy by testing the network with a set of validation data. The accuracy is a function of true and false-positives, if performing classification, and these two metrics are important in desiring a generalized neural network that has not been over-trained. In the case of continuous data, typical R^2 or other metrics may be used to determine accuracy. Ten-fold cross validation is utilized here to ensure that over-fitting does not occur (Kohavi, 1995).

The data that the network is trained with is of the utmost importance. Artificial neural networks are typically referred to as a "black-box" due to the inability to determine what exactly has been conducted as the network is being trained. All that remains is a set of nodes with various weights attached to them that hold no physical/intuitive meaning. As such, the preprocessing of data for a neural network is very important. Contradicting data leads to poor predictions and data with fairly little variability would lead to a lack of generalizability.

One property of the dataset that greatly benefits a neural network is monotonicity, which refers to the fact that the values do not increase and then decrease, or vice versa. This is another way of saying contradicting data. This may not, however, always be satisfied, especially with medical attributes which may be complexly related. Outlying data or highly contradictive data, however, should be removed through some form of analysis of the data. Here, DEA is deployed to perform this function. Preprocessing of the dataset for the ANN is where most of the effort should be focused, for it is here that increases the accuracy and also generalizability is ensured.

3.5.3 DEANN Methodology Implementation

All results in this work were obtained through scripts written in the R language. Freely available *lpSolve* package (under LGPL 2) is utilized for the 'lp' function, which is a linear programming solver. This allows DEA to be implemented and makes it highly modifiable and easily extensible towards new developments. It also results in a fast run-time (considering the number of DMUs considered) and allows consideration of thousands of DMUs simultaneously. In this work, stratification is performed on the entire 12,744 records. The *RSNNS* package by Bergmeir and Benitez (2012) allows use of an 'mlp' function, a multi-layer perceptron network that can be trained by a variety of methods and allows specific choice of number of hidden neurons, learning rate, and other parameters. This package also contains many other useful functions that allow easy normalization, class decoding, etc.

It should be noted that in standard ten-fold cross validation the one-tenth of the dataset reserved for testing is rotated through the entire set. However, in this work, due to stratification, the training data are now scattered throughout the entire dataset. In order to perform ten-fold cross validation those specific records which happen to be in the one-tenth reserved for testing are excluded from the training set and the analysis is continued as normal. This allows specific efficient layers to be utilized for training and the entire dataset to be used for testing while reducing the risk of over-fitting. The dataset was also randomized with respect to the records so that they were unordered to prevent bias within specific folds of the cross-validation.

4. Results and Discussion

To test the accuracy of the ANN a discrete multi-class error metric is used that allows for less error in those predictions which are placed in classes nearer to the actual. Receiver operator characteristic (ROC) plots, used to show sensitivity and specificity, are not utilized here as FUNC_STAT_TRF is not a true class variable, merely a discretized continuous variable representing a percent of functional status (10%, 20%, …, 100%). A discrete accuracy metric which is developed in this study is presented as in Eq. (8):

$$\text{Multi}-\text{Class Accuracy} = \frac{1}{n}\sum\left(1-\frac{\left|x-\hat{x}\right|}{|x_{max}-x_{\min}|}\right) \tag{8}$$

where n is the number of records, x is the actual value, $\hat{x}$ is the predicted value, and x_{max}, x_{min} are the maximum and minimum of the actual values, respectively.

The DEA stratification model was run on the entire subset utilized for this study, yielding twelve levels of efficiency. Table 3 is the mean and variance of the FUNC_STAT_TRF variable of each individual level. As the level number increases, the mean decreases, which means the patient status is decreasing, with the largest variance for the middle levels, at which there may be many efficient records both in terms of high and low patient status. This shows that DEA is separating, by levels, high and low patient statuses based on their inputs. This can also be seen in Table 4, which contains the correlations of single levels.

Table 3. Mean and variance of individual efficiency levels from DEA

Level	Mean	Variance
1	92.63158	245.27
2	86.81111	441.41
3	83.21292	502.34
4	79.82329	501.37
5	77.73079	479.53
6	70.73362	560.30
7	56.95415	715.85
8	51.34238	475.69
9	47.42647	418.67
10	50.10989	293.74
11	27.51634	187.21
12	25.42169	71.47

Table 4. Correlation of FUNC_STAT_TRF with other variables for individual efficiency levels

Level	BMI_TRR	BMI_DON	AGE	AGE_DON	DAYSWAIT_CHRON	FUNC_STAT_TRR	GTIME
1	0.08	0.25	0.09	0.13	-0.10	-0.10	-0.33
2	0.14	0.11	-0.11	-0.03	-0.01	0.10	-0.29
3	0.07	0.16	0.07	0.04	-0.18	0.20	-0.36
4	0.18	0.11	0.17	0.10	-0.11	0.14	-0.31
5	0.19	0.17	0.14	0.20	0.01	0.10	-0.26
6	0.21	0.20	0.14	0.13	0.04	0.27	-0.04
7	0.01	0.15	0.10	0.30	0.02	0.07	0.08
8	0.00	0.00	0.05	0.23	0.01	0.11	-0.14
9	-0.01	0.13	0.06	0.04	0.10	-0.09	-0.33
10	0.39	0.40	0.32	0.42	0.10	0.01	-0.16
11	0.19	0.25	0.26	0.27	0.20	0.37	-0.33
12	0.09	-0.13	0.27	0.14	0.10	-0.03	-0.51

Table 4 shows that individual levels still result in patient status being significantly correlated with one or more inputs. Which variable it is correlated with changes due to each layer containing different patient statuses, as discussed in Table 3 earlier. Level 12 effectively contains only very low patient statuses and is positively correlated with AGE and AGE_DON, which shows that younger age may not necessarily result in a better patient status (recall that inputs were inverted as they were tapped into DEA). The number of days the recipient has waited is negatively correlated with patient status in the first few levels and as the levels increase, it becomes positively correlated, which makes medical intuitive sense considering that a longer time spent on the waiting list should have a negative impact. Another interesting discovery of this study to note is that BMI of the donor and the recipient and FUNC_STAT_TRR (i.e. the functional status at time of transplant) are not strictly positively (or negatively) related to the functional status of the patient at follow up. GTIME is also negatively correlated with recipient status for all levels. This may be due to the fact that DEA only considers a limited number of variables compared to the many available, but it still shows an interesting relationship between the two.

One other set of interesting DEA stratification results is tabulated as in Table 5, which is the correlation of the recipient status with level sums. The results here illustrate how correlation of recipient status effectively decreases as records of individual levels are added together. Level sum 12, for example, is the sum of all efficiency levels and would return the original dataset. The correlation of recipient status with all variables is very low, unlike for level 1, which considers only a single level, and has a significant correlation with many inputs. This high correlation for individual levels is what would hypothetically be highly beneficial to machine learning processes. Training with uncorrelated data results in contradictory observations tapped into the prediction method which may yield low accuracy. Consideration of fewer efficiency levels reduces those contradictions.

Table 5. Correlation of FUNC_STAT_TRF with other variables for sums of efficiency levels

Level Sums	BMI_TRR	BMI_DON	AGE	AGE_DON	DAYSWAIT_CHRON	FUNC_STAT_TRR	GTIME
1	0.08	0.25	0.09	0.13	-0.10	-0.10	-0.33
2	0.11	0.13	-0.04	0.01	-0.04	0.00	-0.29
3	0.06	0.12	0.01	0.03	-0.12	0.07	-0.31
4	0.07	0.09	0.05	0.03	-0.11	0.08	-0.29
5	0.07	0.09	0.06	0.05	-0.06	0.06	-0.27
6	0.07	0.08	0.05	0.03	-0.06	0.07	-0.22
7	0.06	0.06	0.04	0.03	-0.05	0.03	-0.14
8	0.04	0.04	0.04	0.03	-0.05	0.03	-0.06
9	0.04	0.04	0.04	0.02	-0.04	0.03	-0.01
10	0.04	0.03	0.03	0.01	-0.04	0.03	0.02
11	0.04	0.03	0.02	0.00	-0.04	0.02	0.06
12	0.04	0.03	0.02	-0.01	-0.04	0.02	0.08

The ANN was configured to use back-propagation algorithm for weight determination with a hidden layer of twenty neurons. The ANN was then trained on a sum of levels, such as the first, first plus second, so on and so forth up until the sum of all levels. Level one contains 1045 records, levels one and two comprise 2845, and so on until the combination of all levels yields the original subset.

The results of the ANN testing can be seen in Tables 6 through 8. Table 6 represents the confusion matrix for the testing data of chosen level sums, while Tables 7 and 8 are for levels 1-3 and levels 1-12, respectively, 12 being the maximum number of levels.

Table 6. Confusion Matrix for Level One, First Fold.

Actual \ Predicted	5	6	7	8	9	10
1	0	5	0	3	5	51
2	0	1	3	1	0	31
3	0	0	3	2	1	34
4	0	0	2	0	0	32
5	2	0	3	0	1	48
6	0	5	7	3	12	107
7	0	0	5	2	0	101
8	0	1	4	15	1	181
9	0	3	5	2	12	192
10	0	3	8	7	3	367

Table 7. Confusion Matrix for Levels One-Three, First Fold.

Actual \ Predicted	2	4	5	6	7	8	9	10
1	3	0	0	0	0	3	21	37
2	4	0	1	0	0	2	12	17
3	0	1	0	1	1	4	3	30
4	0	7	1	0	0	2	8	16
5	0	0	7	1	3	2	7	34
6	0	1	0	6	3	9	26	89
7	0	2	2	0	2	13	22	67
8	1	1	0	1	3	70	40	86
9	0	0	1	0	2	12	118	81
10	0	0	0	9	0	14	19	346

Table 8. Confusion Matrix for Levels One-Twelve, First Fold.

Actual \ Predicted	2	3	4	5	6	7	8	9	10
1	1	1	0	0	8	4	11	13	26
2	4	0	0	0	3	6	7	3	13
3	3	3	0	0	4	2	11	6	11
4	0	1	2	0	1	4	8	3	15
5	1	3	0	7	9	9	4	5	16
6	3	2	1	1	40	16	16	14	41
7	5	2	1	1	10	34	15	10	30
8	0	3	0	0	3	9	120	14	53
9	0	2	1	0	10	4	7	146	44
10	2	2	3	3	14	20	26	40	278

The confusion matrices show that DEA reduces the generalizability of the dataset. Many more records are predicted as a class 10 due to the way in which there are so many records that have a functional status of ten. Using a single level results in almost no predictions as anything other than a 10 due to the fact that majority of records have a functional status of 10. As the level sums are increased, the ANN regains the ability to generalize and the sum of the first three levels shows that predictions are again appearing on the diagonal. In fact, this by itself signifies the power of the proposed integrated DEANN methodology. There are still great many more predictions as a class 10 compared to the actual, especially for lower values of the functional status, which DEA has most likely thrown out as inefficient due to its lack of ability to handle many of the variables that the ANN considers. Application in reduction of this oversampled class through random sampling or other methods designed to equalize class distributions might prove beneficial towards improving overall accuracy as well as detectability of rarer events.

Using Eq. (8), Table 9 contains the accuracy for the ten folds of cross-validation as well as the mean for the same levels as was displayed in Tables 6-8. It can be seen that in the worst case scenario, utilizing DEA with only level one, the accuracy is reduced from 81.6% to 73.9%. A significant decrease and therefore, the correct choice of level sum is further investigated next in Figure 3.

Table 9. Accuracy for Ten-Folds for Levels One, One-Three, and One-Twelve.

Fold	Level 1	Levels 1-3	Levels 1-12
1	0.742805	0.781266	0.811879
2	0.730682	0.777778	0.811966
3	0.739752	0.772458	0.815542
4	0.733211	0.773155	0.816588
5	0.741148	0.783098	0.829758
6	0.744724	0.789116	0.825135
7	0.736525	0.780045	0.805425
8	0.736264	0.786325	0.821821
9	0.740799	0.779260	0.809262
10	0.744811	0.782138	0.813623
Mean	**0.739072**	**0.780464**	**0.816100**

Figure 3 summarizes the mean accuracy and record count for the training of the ANN at different level sums. In this figure, the record count is represented as a percent of total, with 100% representing the full 12,744 records. The accuracy is fairly stable until level sum 6, at which each decrement in level sum reduces the record count by a significant amount, also decreasing the accuracy. The difference in accuracy can largely be ignored for level sums 9 through 12 due to slight variances when training the ANN. It can be seen that for all level sums the accuracy remains fairly stable, with a maximum difference of roughly 8%. In either case, DEANN method proves superior and increases the performance of prediction for such a large dataset.

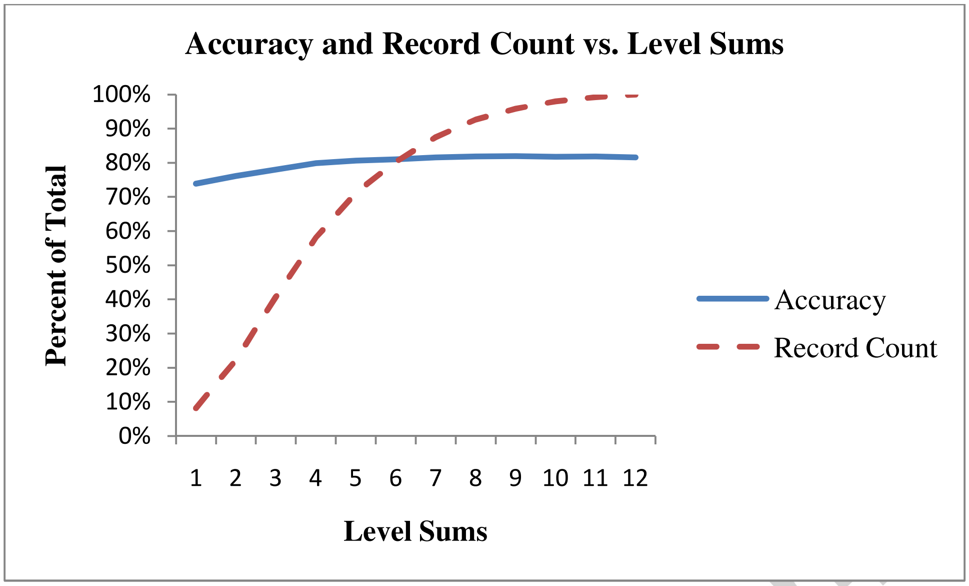

Fig. 3. Accuracy and Record Count Totals as a Function of Efficiency Level Sums

5. Conclusion

The ANN provides a good baseline for the predictions of the functional status of patients, providing acceptable accuracy considering the complex relationship amongst the variables and the high volume of records. On the other hand, DEA efficiency levels effectively separate records based on correlation of inputs and outputs. Individual levels have varying levels of significant positive and negative correlation between inputs and outputs compared to the original dataset which has very low correlation. Consideration of specific level sums maintains an accuracy metric while at the same time reducing the size of the dataset being considered by the ANN which, among other things, reduces the training time of the ANN considerably. This study presents a hybrid methodology, i.e. DEANN, that integrates these two data analytic methods and collectively utilizes the abovementioned features of the two. The viability of the complementary nature of ANN and DEA is presented in this study along with a complex, large, US-based nation-wide healthcare dataset. Although the proposed DEANN method is validated here via a healthcare-based dataset due to its recent popularity in literature, the generic nature of the method renders it viable and practically applicable to other settings that deploy large datasets in a similar fashion. It would hypothetically provide more efficiency in computation of prediction and would be an effective way to deal with such voluminous datasets.

Since conventional DEA approaches work best with continuous data, this current study has utilized only continuous variables available in the UNOS database. However, much research has been done in DEA recently to consider other variable types and it is the authors' intention that this work will be expanded to consider both ordinal and binary data types to better classify the transplants at another study. Future research directions also include the implementation of a modified DEA which considers ordinal and binary values, large scale pruning of the ANN, and reduction of oversampled outputs to further improve training of the ANN. Nevertheless, this study itself provides a strongly acceptable baseline for which these future research goals would improve upon.

References

Abbass, H., 2002. An evolutionary artificial neural networks approach for breast cancer diagnosis, Artificial Intelligence in Medicine 25, 265–281.

Arel, I., Rose, D.C. (2010). Deep Machine Learning – A New Frontier in Artificial Intelligence Research. Computational Intelligence Magazine, IEEE, 5 (4), 13-18.

Athanassopoulos, A., Curram, S., 1996. A Comparison of Data Envelopment Analysis and Artificial Neural Networks as Tools for Assessing the Efficiency of Decision Making Units. Journal of the Operational Research Society 47.

Chen, I.F., 2010. Evaluate the Performance of Cardholders' Repayment Behaviors Using Artificial Neural Networks and Data Envelopment Analysis. Networked Computing and Advanced Information Management 478–483.

Chen, Y., 2005. Measuring super-efficiency in DEA in the presence of infeasibility. European Journal of Operational Research 161, 545–551.

Christoph Bergmeir, Jose M. Benitez (2012). Neural Networks in R Using the Stuttgart Neural Network Simulator: RSNNS. Journal of Statistical Software, 46 (7), 1-26. http://www.jstatsoft.org/v46/i07/.

Cook, W., Tone, K., Zhu, J., 2014. Data envelopment analysis: Prior to choosing a model. OMEGA, Vol. 44, 1-4.

Cook, W.D., and Zhu, J., Rank order data in DEA: A general framework, European Journal of Operational Research, Vol. 174, Issue 2 (2006), 1021-1038.

Dayhoff, J.E., DeLeo, J.M., 2001. Artificial Neural Networks - Opening the Black Box. Cancer Supplement 91 (8).

Delen, D., Oztekin, A., Kong, Z.J., 2010. A machine learning-based approach to prognostic analysis of thoracic transplantations. Artificial Intelligence in Medicine 49(1), 33-42.

Dvorchik, I., Subotin, M., Marsh W., McMichael J., Fun J., 1996. Performance of multi-layer feedforward neural networks to predict liver transplantation outcome. Methods of Information in Medicine 35, 12–18.

Dybowski, R., Gant V., 2001. Clinical Applications of Artificial Neural Networks, Cambridge University Press.

Hatami-Marbini, A., Emrouznejad, A., Tavana, M., 2011. A taxonomy and review of the fuzzy data envelopment analysis literature: Two decades in the making. European Journal of Operational Research 214 457-472.

Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of the 14th International Conference on AI (IJCAI) (pp. 1137–1145). San Mateo, CA: Morgan Kaufmann.

Kotsiantis, S., Kanellopoulos, D., Pintelas, P., 2006. Data Preprocessing for Supervised Learning. International Journal of Computer Science 1 (2).

Kuo, R., Wang, Y., Tien, F., 2010. Integration of artificial neural network and MADA methods for green supplier selection. Journal of Cleaner Production 18, 1161-1170.

Liu, J., Lu, L., Lu, W.M., Lin, B., 2013. Data envelopment analysis 1978-2010: A citation-based literature survey. OMEGA, Vol. 41, Issue 1, 5-15.

Liu, J., Li, L., Fu, C., Wu, Z., 2009. A Military Maneuver Engineering Support Evaluation Model Based on ANN and Super-efficiency DEA. Intelligent Systems and Applications 1–4.

Matin, R., Kuosmanen, T., 2009. Theory of integer-valued data envelopment analysis under alternative returns to scale axioms. OMEGA, Vol. 37, Issue 5, 988-995.

Meisel, S., Mattfeld, D., 2010. Synergies of Operations Research and Data Mining, European Journal of Operational Research 206, 1-10.

Nahra, T., Mendez, D., Alexander, J., 2009. Employing super-efficiency analysis as an alternative to DEA: An application in outpatient substance abuse treatment. European Journal of Operational Research 196, 1097-1106.

Olanrewaju, O., Jimoh, A., Kholopan, P., 2012. Integrated IDA-ANN-DEA for assessment and optimization of energy consumption in industrial sectors. Energy 46, 629–635.

Olanrewaju, O., Jimoh, A., Kholopane, P., 2011 Energy Efficiency Assessment Using Artificial Neural Network Combined with Data Envelopment Analysis. IEEE Africon, 13–15.

Ozdemir, D., Temur, G.T., 2009. DEA ANN Approach in Supplier Evaluation System. World Academy of Science, Engineering and Technology 54, 343–348.

Oztekin, A., Delen, D., Kong, Z.J., 2009. Predicting the graft survival for heart–lung transplantation patients: An integrated data mining methodology. International Journal of Medical Informatics 78, 84-96.

Oztekin, A., Kong, Z., Delen, D., 2011. Development of a structural equation modeling-based decision tree methodology for the analysis of lung transplantations. Decision Support Systems 51, 155-166.

Pendharkar, P.C., 2005. A Data Envelopment Analysis-Based Approach for Data Preprocessing. IEEE Transactions on Knowledge and Data Engineering 17 (10).

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.

Saljooghi, F.H., Giski, N.E., 2013. Qualitative factors in data envelopment analysis. Annals of Fuzzy Mathematics and Informatics 5 (2), 387–397.

Samoilenko, S., Osei-Bryson, K.M., 2013. Using Data Envelopment Analysis (DEA) for monitoring efficiency-based performance of productivity-driven organizations: Design and implementation of a decision support system. OMEGA, Vol. 41, Issue 1, 131-142.

Samoilenko, S., Osei-Bryson, K.M., 2010. Determining sources of relative inefficiency in hetereogeneous samples: Methodology using Cluster Analysis, DEA and Neural Networks. European Journal of Operational Research 206, 479-487.

Seiford, L.M., Zhu, J., 2003. Context-dependent data envelopment analysis-- measuring attractiveness and progress. OMEGA, Vol. 31, Issue 5, 397-480.

U.S. Renal Data System Coordinating Center (2013), USRDS 2013 Researcher's Guide to the USRDS Database.

Wang, S., 1995. The Unpredictability of Standard Back Propagation Neural Networks in Classification Applications. Management Science 41 (3).

Wang, S., 2003. Adaptive non-parametric efficiency frontier analysis: a neural-network-based model. Computers and Operations Research 30.

Zhu, J., 2014. Quantitative Models for Performance Evaluation and Benchmarking: Data Envelopment Analysis with Spreadsheets. 3rd edition. Springer, New York.

Zhu, J., 2001. Super-efficiency and dea sensitivity analysis. European Journal of Operational Research 129, 443–455.

Zhuang, Z., Churilov, L., Burstein, F., Sikaris, K., 2009. Combining data mining and case-based reasoning for intelligent decision support for pathology ordering by general practitioners, European Journal of Operational Research 195, 662-675.

Highlights:

- A hybrid methodology, DEANN, for prediction improvement was developed.
- DEA was utilized to classify the dataset in efficiency frontiers.
- Predictions were performed using an ANN due to the complexity of dataset.
- Implementation was performed on an organ transplant dataset.
- High accuracy rates with a reduction in training dataset size validate the DEANN.
- This generic approach is readily applicable to a wide number of areas.