Efficiency Measurement and Improvement Projection of Container Terminals

Song Jin Yu · Mithun. J. Sharma

Division of Shipping Management, Korea Martime Uiversity
coppers@bada.hhy.ac.kr, mithun sharma@hotmail.com

Abstract

Data Envelopment Analysis (DEA) is a multifactor productivity measurement tool and is used in assessing the relative efficiency of homogenous units. DEA assumes that the decision making unit (DMU) are homogenous in their environment and avoids any error or noise in measurements. Container terminals, which act as an interface between the sea and the shore, for loading and unloading of containers from ship to shore andvice-versa, may operate with its own attributes and goals. Every container terminal is characterized by some physical values that represent different relevant properties of the terminal. DEA, if employed alone, to measure the efficiency and set the bench mark for inefficient terminals gives biased result because all the container terminals may not be inherently similar. In order to overcome this shortcoming, in this paper, two important fields of information technology: data mining and data envelopment analysis is integrated to provide a new tool to appropriately set bench mark for inefficient terminals and prioritize the technical inputs that have the greatest impact needed to improve the inefficient terminals which otherwise is not possible with DEA alone.

Key words : Data Envelopment Analysis, decision making unit, data mining, bench mark.

1. Introduction

The rapid growth of containerization around the world has brought a significant redefinition in the ports and shipping sector. Containerization, the movement of cargo in containers, is a system with an ocean component and a land component. A container terminal is a facility which provide a package of activities and services to handle and control container flows from vessel to rail, or road, and viceversa. The container terminal is the interface between the ocean and land modes of transport and a major component of containerization system. The latter is a dynamic system within which various enterprises (carriers, terminal operators, stevedores, labour, port authorities, shippers, railways, truckers, government and others) interact. Each influences productivity and at one time or another may be the primary determinant or constraint on control of productivity at a specific terminal or within the entire system.

There are various papers based on efficiency measurement of container port industry in relation to productive activities: e.g.Cullinane etal 2004-05; Tongzon 2001. In particular, non-parametric frontier methods Data Envelopment Analysis (DEA)have been developed with application across a wide range of sectors. In general, given a set of already established and operating Decision Making Units, DMUs with corresponding inputs and outputs data DEA persecutes two main goals; (a) It divides the existing DMUs into efficient and inefficient units and gives each DMU a score of efficiency. An efficient DMU usually obtains a score 1, (b) for every inefficient DMU0 it attempts to find an efficient reference point, i.e. a point on the envelopment surface that DMU0 can consider an achievement point.

The way to achieve this goal is not unique resulting in several DEA models. The choice of the set of efficient DMUs affects the geometry of the envelopment surface. It also determines the returns of scale (if any)the models work with. The efficient reference point is given by the projection of the given DMU on the envelopment surface. The choice of such projection also determines the efficiency score. The main requirement of a benchmarking is for the improvement projection of an inefficient unit.

Although benchmarking in DEA allows for the identification of targets for improvements, it has certain limitations. A difficulty addressed in the literature regarding this process is that an inefficient DMU and its benchmarks may not be inherently similar in the operating practices. Doyle and Green, 1994; Talluri and Sarkis, 1997, have used clustering based method to overcome this problem. Thanassoulis,1995, grouped the units by the characteristics of input resource mixbut not according to their efficiency levels. In order to overcome this problem, in this paper, DEA recursive analysis is used at first to segregate the units based on the efficiency score and then unsupervised clustering tool KSOM and a decision tree is employed to cluster them with similar input properties and discriminate the input attributes so as to make a feasible decision for stepwise improvement at the respective segregated efficiency levels.

This paper provides a hybrid methodology in order to overcome the shortcomings of Data Envelopment Analysis, DEA which is a multi-factor productivity analysis tool used in assessing the technical inputs of container terminals. The objectives are summarized as follows: (a) DEA recursive analysis is used to segment the container terminals based on their efficiency score, (b) the container terminals with similar properties are clustered using unsupervised clustering method Kohnen's Self-Organizing Map, KSOM that can be utilized as benchmarks for improvement, and (c) the input attributes of container terminals are discriminated based on their grades using a discriminant descriptor See 5, a decision tree analyzer, which gives priority to the variables required for the improvement of a container terminal.

2. Literature Review of Data Envelopment Analysis

DEA is a multi-factor productivity analysis model for measuring the relative efficiencies of a homogenous set of decision making units (DMUs). The efficiency score in presence of multiple input and output factors is defined as:

$$\text{Efficiency}=\frac{\text{Weighted sum of outputs}}{\text{Weighted sum of inputs}} \qquad (1)$$

Assuming that there are n DMUs, each with m inputs and soutputs, the relative efficiency score of a test DMU p is obtained by solving the following model proposed by Charnes et al. (1978):

$$\text{Max } h_0 = \frac{\sum_{r=1}^{s} U_r Y_{r0}}{\sum_{r=1}^{m} V_i Y_{i0}} \qquad (2)$$

$$s.t \quad \frac{\sum_{r=1}^{s} U_r Y_{rj}}{\sum_{i=1}^{m} V_{is} X_{ij}} \leq 1$$

$$U_r, Y_i \geq 0$$

where
U_r = weight given to output r
V_i = weight given to input i
X_{ij} = amount of input i utilized by DMU j
Y_{rj} = amount of output r utilized by DMU j
s = the number of input variables
m = the number of output variables
n = the number of DMU

The fractional problem shown as (2) can be converted to a linear program as shown in (3)

$$\text{Max } h_0 = U_r Y_{r0} \qquad (3)$$
$$s.t. \quad V_i X_{i0} = 1$$

$$\sum_{\gamma=1}^{s} U_r Y_{rj} - \sum_{\gamma=1}^{m} V_i X_{ij} \leq 0$$

$$U_r V_i$$
$$j=1,2,\mathrm{K},n$$

The above problem is run n times in identifying the relative efficiency scores of all the DMUs. Each DMU selects input and output weights that maximize its efficiency score. In general, a DMU is considered to be efficient if it obtains a score of 1 and a score of less than 1 implies that it is inefficient.

Container ports can be defined as places with facilities for shipping lines where equipments are available to handle container flow from vessels to rail or road and vice versa. The subject of container port performance measurement is an important issue facing port management. A model, or models, that directly linked the inputs or factors of production at each container port - labour, capital equipment and land - to its outputs could give some insights into productivity performance(Modern ports: A UK policy).

In the earlier years a common feature in measuring the port performance is the use of partial indicators such as use of capacity, number of labors, waiting time etc. These partial indicators are all useful but they can be quite misleading since they do not necessarily generate the same ranking of ports (Antonio et al 2001).

The second generation of studies relying on formal measures of efficiency is an attempt to address this failure. In general, researchers focus on ports cost or performance measurement to make the most of the information available. Roll and Hayuth (1993) rely on data commonly available from annual reports in ports and Tongzon (2001) covers 16 ports for which he obtained comparable data for 1996.The model preferences are evenly distributed between stochastic frontier and Data Envelopment Analysis, DEA, where Liu (1995) focuses on production to calculate technical efficiency and compares the influence of public and private ownership in Britain. Roll and Hayuth (1993) show how DEA can be useful in assessing the relative effectiveness of various ways of organizing port services when limited data is available. Martinez, Diaz, Navarro and Ravelo (1999) rely on a DEA to assess the relative efficiency of Spain's ports. Tongzon (2001) uses DEA to make an international comparison of efficiency of 4 Australian and 12 other ports from around the world.

Furthermore, Cullinane, Song, Wang applies DEA windows analysis, utilizing panel data, to a sample of the world's major container ports in order to deduce their relative efficiency. Their results show that panel data prevails over cross sectional data.

Stochastic frontier methods have also been applied tocontainer port industry where measures of physical quantities of merchandises have been adopted by Martinez et al (1999), Roll and Hayuth (1993). Liu (1995) and Coto et al (2000) assume a single output technology and measure output through the volume of merchandise handled. For two papers with cost functions, Banos et al (1999) and Coto et al. (2000), labor prices are approximated by the ratio of total labor cost to the number of workers and the price of capital is obtained by dividing the amortization of the period by the length of docks in Coto et al, 2000. Table2.1 gives a survey of literature on efficiency measures in the port sector.

Table2.1 Survey of the literature on efficiency measures in the ports sector

Author	Data(1)	Model(2)
Liu (1995)	Panel data, 28 UK ports, 1983-1990	SPF
Coto, Banos and Rodriguez (1999)	Panel data, 27 Spanish ports	SCF
Roll and ayuth(1993)	Cross section data, 1993	DEA
Martinez et al(1999)	Panel data, 26 ports, 1993 - 1997	DEA
Tongzon (2001)	Panel data, 16 ports, 1996	DEA
Cullinane, Song and Wang, Ji (2004)	Panel data, 30 ports, 1992 - 1999	DEA Windows nalysis

(1) To indicate sample size
(2) SPF: Stochastic Production Frontier; SCF: Stochastic Cost Frontier; DEA: Data Envelopment Analysis

3. Efficiency Measurement of Container Ports

3.1 Research Design

As discussed in the previous chapter we have relatively modest amount of papers dealing with container port efficiency measurement. Valentine and Gray (2001) compared port efficiency with particular type of ownership or organization, Cullinane et al (2005) examines the relationship between port privatization and relative efficiency within the container port industry. Their paper concludes with the rejection of hypothesis that greater private sector involvement in container port sector irrevocably leads to improved efficiency.

In this paper, 70 container terminals from around the world are taken as sample study based on the argument that container terminals are more suitable for one-to-one comparison than whole container ports (Wang, Song, Cullinane, 2002). Thisstudy is carried out to set favorable benchmarks for inefficient container terminals and step-wise selection of input attributes for improvement at their respective levels.

The DEA yields a detailed analysis for DMUs to determine the efficient and inefficient units in order to gain useful information for making further improvements. The information can discover unknown relationships among the data which includes identifying the most productive operating scale sizes, the saving in resources, and the most suitable ways to enhance inefficient units (Thannassoulis, 2001). DEA alone is inefficient to set appropriate benchmark for container terminals where the input factors are heterogeneous and if assessed by DEA without modification yields biased results. In order to overcome the problem this paper integrates two important fields of information technology: data mining and data envelopment analysis.

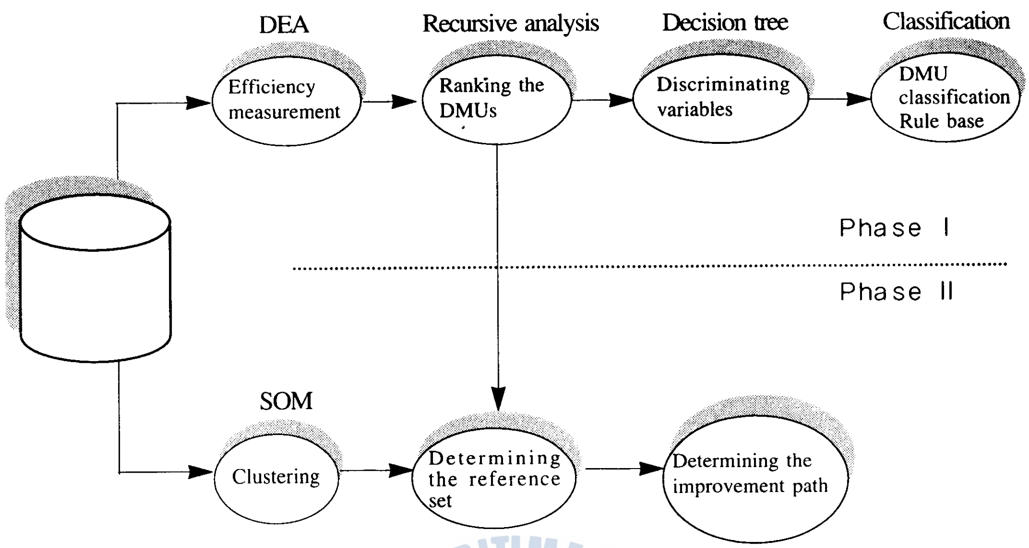

Figure 3.1 Research Framework

The first problem in DEA model is that it assumes that all DMUs are homogenous and identical in operations (SS 94). Since various applications have heterogeneous DMUs and there is a need to evaluate these applications under the DEA due to its acceptance as a performance measurement in different kind of business, we have to modify the DEA to work with these applications. For every inefficient DMU, DEA identifies a set of corresponding efficient units that can be utilized as benchmarks for improvement. The benchmark can be obtained from the dual problem shown below:

$$\text{Min } \theta \qquad (4)$$

$$\sum_{j=1}^{n} \lambda_j X_{ij} - \theta X_{ip} \leq 0 \quad \text{for } i = 1,2,\mathrm{K},m$$

$$\sum_{j=1}^{n} \lambda_j Y_{rj} - Y_{kp} \geq 0 \quad \text{for } r = 1,2,\mathrm{K},s$$

$$\lambda_j \geq 0 \quad \text{for } j = 1,2,\mathrm{K},n$$

where

θ = efficiency score, and

λ_j = dual variables.

Based on the problem (4), a test DMU is inefficient if a composite DMU (linear combination of units in the set) can be identified which utilizes less input than the test DMU while maintaining at least the same output levels. The units involved in the construction of composite

DMU can be utilized as benchmarks for improving the inefficient DMU. DEA also allows for computing the necessary improvements required in the inefficient unit's inputs and outputs to make it efficient. Although benchmarking in DEA allows for the identification of targets for improvements, it has certain limitations. Aninefficient DMU and its benchmarks may not be inherently similar in their operating practices. To overcome these problems researchers have utilized performance based clustering methods for identifying more appropriate benchmark (Doyle and Green 1994; Talluri and Sarkis 1997). These methods cluster inherently similar DMUs into groups, and the best performer in a particular cluster is utilized as a benchmark by other DMU in the same cluster. These studies have proposed clustering method in various DEA applications but after evaluating the efficiency score for DMUs. Based on the result of DEA, they built clusters for each DMU and its reference set to show the degree of sensitivity in the presence of a particular DMU in the cluster. This direction is not suitable in analyzing the non homogenous DMUs due to (a) DEA will overestimate the efficiency scores of those operating under favorable conditions, and (b) DEA will underestimate the efficiency scores of those operating under unfavorable conditions.

The proposed approach integrates unsupervised learning tool Kohnen's Self-Organizing Map, KSOM and a typical decision tree analyzer See5.

The methodology is divided into two phases. Before entering the phases DEA is applied to evaluate the efficiency of container terminals with its multi-dimensional inputs and outputs.

Here, the standard model, DEA-CCR input-oriented is applied to evaluate the overall performance measure of container terminals.

After performing the above analysis the phases are entered eventually.

Phase Ⅰ

Segments the container terminal based on their efficiency score obtained from recursive analysis. The segregated terminals form tiers according to their efficiency level which are labeled accordingly, such as tier 1, 2, 3 etc.

After segmenting the container terminals on the basis of their efficiency score the input attributes are discriminated ateach level based on the information theory.

A typical discriminant descriptor, See5 is used that extracts informative patterns from data.

The attributes are discriminated in order to prioritize them so that decision taken on improvement of technical inputs is feasible. After See 5 is invoked, with default values, it creates a decision tree. The last section of the See5 output concerns the evaluation of the decision tree, first on the cases in container terminal data from which it was constructed, and then on the new cases in test file.

Phase II

Here Kohnen's Self-Organizing Map, which is one of the clustering tools for grouping similar units according to the characteristics of input variables, is utilized. The KSOM is an unsupervised learning tool to classify data. The SOM tool clusters the terminals based on their input characteristics which helps to appropriately benchmark the inefficient terminals for improvement.

3.2 Determining the Technical Inputs

3.2.1 Efficiency Evaluation of Container Terminals Input / output data set

In the most general sense productivity measures output per unit of input. Container terminal productivity deals with the efficient use of labor, equipment and land. Terminal productivity measurement is a means of quantifying the efficiency of the use of these three resources (Dowd and Leschine,1990).Song et al, 2003 discuss that input and output variables should reflect the actual objectives and process of container port production as accurately as possible. The goal of a container terminal determines the definition of variables.

In this paper the main objective is to check which technical inputs affect the efficiency of a container terminal since a container terminal uses equipments that are capital intensive. The maximization of output verily relies on the operational and technical aspect of the terminal. When a ship arrives at the port, quay cranes(QCs) take the import containers off the ship's hold or off the deck. Next, the containers are transferred from the QCs to vehicles that travel between the ship and the stack. This stack consists of a number of lanes, where containers can be stored for a certain period. Equipments, like cranes or straddle carriers (SCs), serve the lanes. A straddle carrier can both transport containers and store them in the stack. It is also possible to use dedicated vehicles to transport containers. If a vehicle arrives at the stack, it puts the load down or the stack crane takes the container off the vehicle and stores it in the stack.

After a certain period the containers are retrieved from the stack by cranes and transported by vehicles to transportation modes likebarges, deep sea ships, trucks or trains. To load export containers onto a ship, these processes are executed in reverse order. Most of the terminals make use of manned equipments, like straddle carriers, reach stackers, cranes and multi-trailer-systems.

The unloading and loading process at a typical container terminal as shown in figure 2

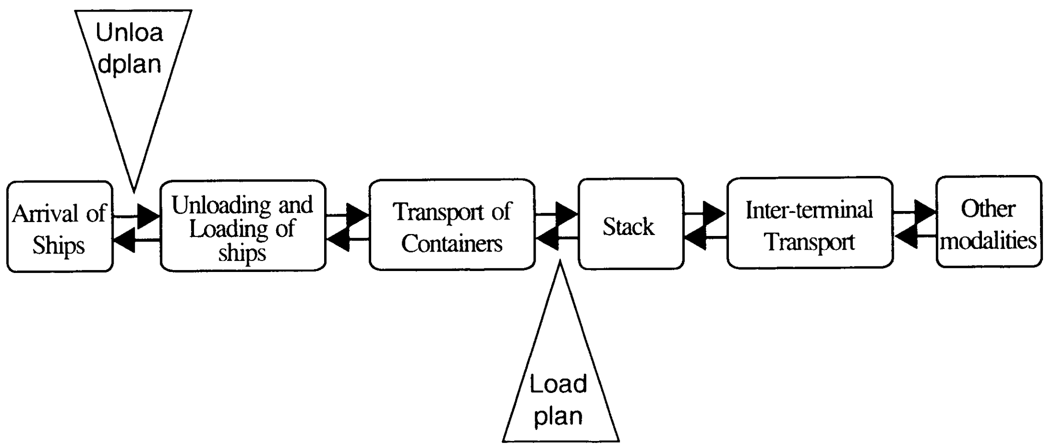

Figure 3.2 Container Process flow chart

Automated and manned terminals both use quay cranes. QCs are manned because automation of this process encounters practical problems, like exact positioning of containers. The QCs have trolleys that can move along the crane arm to transport the container from the ship to the transport vehicle and vice versa. A spreader, a pick up device attached to the trolley, picks the containers. The QCs move on rails to the different holds to take/put containers off/on the deck and holds. It can occur that at the same moment one QC is unloading containers while another QC is loading containers.

Transport of containers from ship to stack and vice versa

For the transport of a container at a manned terminal, vehicles like forklift trucks, reach stackers, yard trucks or straddle carriers can be used.

Stacking of containers

Two ways of storing containers can be distinguished: storing on a chassis and stacking on the ground. With a chassis system each container is individually accessible. With stacking on the ground containers can be piled up, which means that not every container is directly accessible. As a consequence of limited storage space, nowadays stacking on the ground is most common. The stack is the place where import and export containers can be stored for a certain period. The stack is divided into multiple blocks/lanes, each consisting of a number of rows. The height of stacking varies per terminal between two and eight containers high. At the end of each lane a transfer point might be situated. At this point the crane takes places the container off/on the vehicle that transports the container. Empty containers are usually stored separately. A decision that has to be made is choosing the type of material handling equipment that will take care of the storage and retrieval of containers in and from the stack. Systems like forklift trucks, reach stackers, yard cranes and straddle carriers can be chosen. Yard cranes move on rubber tires or on rails over the containers. They can provide high density storage and can be automated. These automated cranes are called Automated Stacking Cranes (ASCs). ASCs move on rails and are controlled by the central operating system. The ASC takes places the container with a spreader from/on the AGV. Based on the processes of container terminal operation the terminal area and quay length are the best input variables for *'land'* factor and number of quay gantry cranes, the number of yard gantry cranes, the number of straddle carriers and the number of reach stackers are best input variables for *'equipment'* factor.

The labor hours in this paper is taken as the input variable for *'labor'* factor due to the availability of data.

Container movements-numbers

As per the output factor is concerned the container throughput is undoubtedly the most important and widely accepted indicator. There are two measuring units in general use that indicates waterfront productivity in terms of the throughput of container;they are 1) Container movements- TEUs and 2) Container movements- numbers. Using the second unit of measure of container movements can overcome to some extent the deficiency that the TEU measure is affected by the mix of twenty-foot and forty-foot containers. This measure simply counts the number of container movements per hour regardless of the size of the containers. Another consideration is that container throughput is the most appropriate and analytically tractable indicator of the effectiveness of the production of a terminal or port. Fig.3.3 shows the systematic organization of input and output variables.

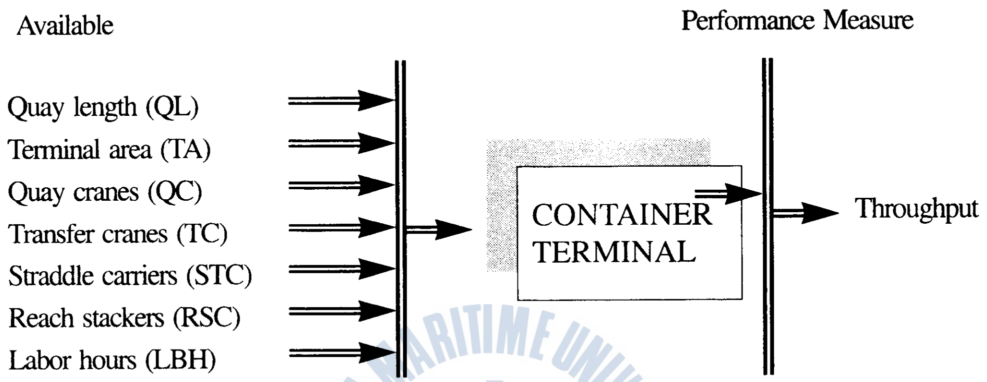

Figure 3.3 Input/output Data Set

Table 3.1 Summary of Variables

Varible		Mesaurement
Input factors	Quay length (QA)	Total quay length of a container terminal
	Terminal area (TA)	Total area of a container terminal
	Quay cranes (QC)	Total number of quay gantry cranes
	Transfer cranes (TC)	Total number of yard cranes (RTG, RMG)
	Straddle carriers (STC)	Total number of straddle carriers
	Reach stackers (RSC)	Total number of stacker vehicles
	Labor Hours (LBH)	Total working hours
Output factor	Throughput	The number of container movement per year

3.2.2 Efficiency Measurement of Container Terminals

Container terminals are more suitable for one-to-one comparison as argued by Wang, Song, Cullinane, 2002. Based on their argument measurement of 70 container terminals from around the world is done by using the DEA. DEA calculates a maximal performance measure for each unit relative to all other units in the observed population with the sole requirement that each DMU lie on or below the frontier. The units which are not on the frontier are scaled against a convex combination of the units on the frontier facet closet to it.

It is important to note that DEA calculations, because they are generated from actual observed data for each DMU, produce onlyrelative efficiency measures. The relative efficiency of each unit is calculated in relation to all other units, using the actual observed values for the outputs and inputs of each DMU. The calculations are designed to maximize the relative efficiency score of each DMU, subject to condition that the set of weight obtained must also be feasible to all other DMUs involved in the calculation. For each inefficient DMU (one that lies below the frontier), DEA identifies the sources and level of inefficiency for each of the inputs and outputs. The level of inefficiency is determined by comparison to a single referent DMU or a convex combination of other referent DMUs located on the efficient frontier that utilize the same level of inputs and produce same or higher level of outputs.

3.2.3 Segmenting the Container Terminals

In the preceding section DEA was utilized to measure the efficiency of 70 container terminals. DEA determines the most productive group of terminals and least productive terminals i.e, the terminals are segmented into efficient or inefficient group by DEA. Previous studies have shown similar clustering method after evaluating the efficiency score of the DMUs. Thanassoulis (1995) clustered the DMUs using DEA by the characteristics of input resource mix. In this paper, DEA is used recursively to segment the terminals as a result of which various tiers are obtained.

In the first step the efficiency score of entire set of units are obtained. The result of the first analysis reveals the most efficient group of DMUs with a score of 1. This group is labeled as 'tier 1'. In the second step again analysis of the remaining DMUs are carried out where a score of 1 is designated to the efficient DMUs and the group is labeled as 'tier 2'.Similarly, the procedure is repeated until the number of remaining DMUs is at least three times greater (8*3=24) than that of inputs along with outputs, as proposed by Banker et al. (1984). Thus after the recursive analysis a set of tiers with their specific grade is derived as shown in figure 3.4

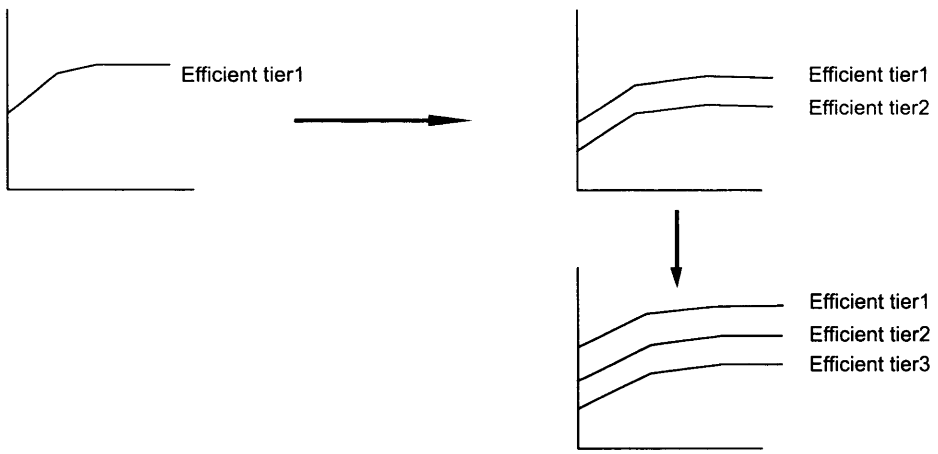

Figure3.4 Tiers based on efficiency

Figure 3.4 shows that the units of tier 1 are superior in efficiency to those in tier 2 and 3.

These segregated DMUs which are based on efficiency obtained from recursiveanalysis is used as input data for a typical decision tree analyzer See 5 to generate classification rule and use them to determine the stepwise improvement path for the inefficient DMUs.

3.2.4 Generating classification rules for each tier using Discriminant Descriptor See 5

Decision tree learning is a method for approximating discrete-valued target functions, in which the learned function is represented by decision tree. Decision trees classify instances by sorting them down the tree from the root to some leaf node, which provides the classification of the instance. Each node in the tree specifies a test of some attribute of the instance, and each branch descending from that node corresponds to one of the possible values for this attribute. An instance is classified by starting at the root node of the tree, testing the attribute specified by this node, then moving down the tree branch corresponding to the value of the attributes. This process is then repeated for the sub tree rooted at the new node.

In this paper, a typical decision learning system, See 5 is used to generate the rule set for classifying the container terminals which adopts a supervised learning scheme that constructs decision trees from a set of examples. The method first chooses a subset of the training examples (window) to form a decision tree. If the tree does not give the correct answers for all the objects, a selection of the exceptions (incorrectly classified examples) is added to the window and the process continues until the correct decision set is found. The eventual outcome is a tree in which leaf carries a class name, and each interior node specifies an attribute with a branch corresponding to each possible value of that attribute.

See 5 uses an information theory approach aiming at minimizing the expected number of tests to classify an object. The attribute selection part of See 5 is based on the assumption that the complexity of the decision tree is strongly related to the amount of information. Information based heuristic selects the attribute providing the highest information gain due to a proposed split to the information gain attributable solely to the number of subsets created as the criterion for evaluating proposed splits.

The system uses information gain ratio as evaluation function for classification, with the following equation (J. Ross Quinlan, 1993),

$$\text{Gain ratio}(X) = \text{gain}(X) / \text{split info}(X)$$

Where split info (X) = $-\sum_{i=1}^{n} |T_i|/|T| * \log 2\,[\,|T_i|/|T|\,]$, gain ($X$) = info ($T$) - info x ($T$) and gain (X) measures the information that is gained by partitioning T in accordance with the test X.

In this work, rules are generated for classifying new units in each tier to determine the input and output variables that will discriminate between the tiers by the degree of affecting the efficiencies of the DMUs (discriminant descriptor).

3.3 Benchmarks for Inefficient Container Terminals

In the second phase of the analysis, an unsupervised clustering tool Kohnen's Self-Organizing

Map is used which clusters the terminals based on their features, for appropriate selection of the efficient peers.

3.3.1 Clustering the Container Terminals using SOM

For every inefficient DMU, DEA identifies a set of corresponding efficient units that can be utilized as benchmarks for improvement. Although benchmarking in DEA allows for the identification of targets for improvements, it has certain limitations. An inefficient DMU and its benchmarks may not be inherently similar in its operating practices, it is here that integration of another tool becomes inevitable. Hence, SOM is employed which groups similar terminals according to the characteristics of the inputs, for the inefficient terminal to select appropriate benchmark for improvement.

The SOMuses unsupervised learning scheme to train neural network (Sabrina Sesito, Tharam S.Dillon, 1994) (Michael et al.1997). Unsupervised learning comprises of those techniques for which the resulting or desired outputs for the training sequences are not known. The network is only told the input vectors, and the network self-organizes these inputs into categories.

Each link between a node in the input layer and a node in the output layer has an associated weight. The net input into each node in the output layeris equal to the weighted sum of the inputs. Learning proceeds by modifying these weights from an assumed initial distribution with the presentation of each input pattern vector. This process identifies groups of nodes in the output layer that are close toeach other and respond in a similar manner. A particular group of units together forms an output cluster. The topology preserving the mapping from the inputs to the cluster reflect the existing similarities in the inputs and capture any regularities and statistical features, and model the probability distributions which are present in the input data.

The SOM uses competitive learning. When an input pattern is imposed on the network, one output node is selected from among all the output nodes as having the smallest Euclidean distance between the presented input pattern vector and its weight vector. This output unit is declared the winner in the competition among all other neurons in the output layer. Only the winning neuron generates an output signal from the output layer. All the other neurons have a zero output signal.

The input weight vectors are usually normalized in a SOM so that they have values between 0 and 1. If the dot products between the normalized input vector X^ and a normalized set of weight vectors W^j are determined, the neuron with the largest dot product (the one with the smallest Euclidean distance) is declared to be the winner. Thus the winner is the vector obtained from the expression:

$$\underset{J}{\mathrm{Max}}(\mathrm{X}^{t} \cdot \mathrm{W}^{t})$$

As learning involves adjustment of weight vectors, learning with this particular input pattern is restricted to lateral interconnections with nearest neighboring units of the winning neuron in the output layer. Adjusting their weights closer to the input vector carries out learning for the nodes within theneighborhood. The size of the neighborhood is initially chosen to be large enough to include all units in the output layer. However, as learning proceeds, the size of the neighborhood is progressively reduced to a pre-defined limit. Thus during these stages, fewer neurons have their weights adjusted closer to input vector. Lateral inhibition of weight vector that are distant from particular input pattern may also be carried out.

A summary of general algorithm for SOM

1. Initialize weights to small random values and set the initial neighborhood to be large. One approach is to set each weight vector equal to an input vector pattern when there are more training input patterns than output units. This approach performs best with very large network and training set.
2. Stimulate the net with a given input vector.
3. Calculate Euclidean distance between the input and the output node and select the output with the minimum distance.
4. Update weights for the selected node and the nodes within the neighborhood.
5. Repetition of these steps from 2 until a stopping criterion is met.

3.3.2 Setting the Benchmark for inefficient Terminals

The result of DEA recursive analysis segregate the terminals based on their efficiency grades i. e efficient terminals in the upper tier becomes a reference set of inefficient terminals in the lower tier. Utilization of SOM helps us cluster the terminals based on their input features. Therefore, SOM is a pattern recognition tool used for clustering and DEA recursive analysis is used for segregating the terminals.

3.3.3 Determining the Improvement Path

The benchmarks derived from DEA recursive analysis is utilized after SOM clusters the terminals. The decision as to which technical input is to be taken into account for improvement at the respective levelcan be derived from the decision tree.

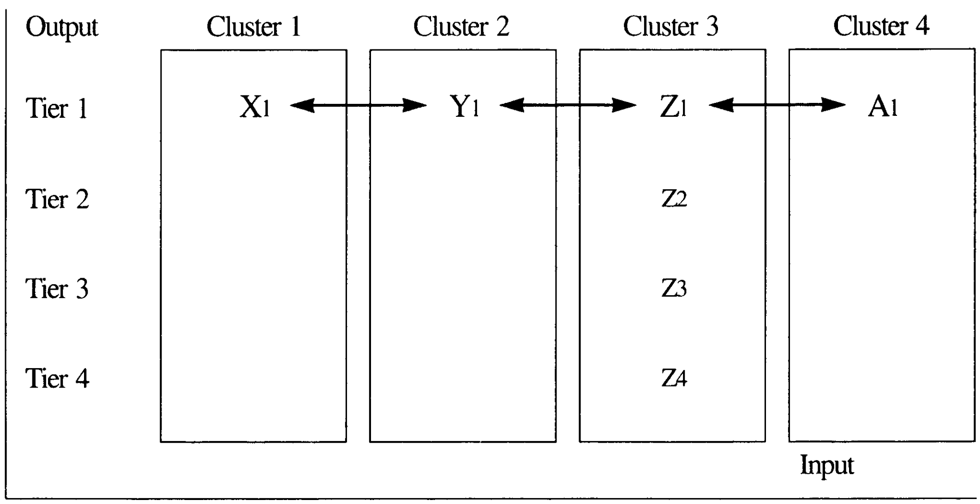

Figure 3.6 A Schematic representations for Improvement Projection
X1, Y1, Z1, A1 --- Container terminals of different clusters but same efficiency score.
Z1, Z2, Z3, Z4 ---- Container terminals of same cluster but different efficiency score.

4. Data Analysis

4.1 Evaluating the Efficiency of Container Terminals using DEA

Data was collectedfrom 70 container terminals from relevant data sources like Containerization International Year Book 2005, The Drewry Annual Container Market Review and Forecast 2004-05 and from specific websites of the port authorities. Here, the DEA-CCR-input model was applied to identify the relative efficiency of container terminals (refer to table 4.1).

The analysis shows that out of the 70 container terminals 19 container terminals are indicated by DEA productivity rating of 100 percent. From the analysis it is seen that some container terminals used an excess of only 4.23% of resources, for example Aarahus container terminal which is close to the efficiency frontier with a score of 95.77%. There are container terminal with score of around 50% indicating the excess use of half of the resources and terminals with 11-12.3% are highly inefficient wasting a significant amount of resources. These findings indicate that some of the container terminals have to make substantial productivity improvement to come up to the highest level that is 100 percent score.

The Efficiency Measurement System (EMS) tool was applied to measure the performance of the container terminals. EMS uses the idea of R.D. Banker and R.C. Morey (1986), Efficiency Analysis for Exogenously Fixed Inputs and Outputs, OR, 34, 513-521. The input oriented measure used in the analysis for the terminals quantifies the input reduction which is necessary to become efficient holding the outputs constant.

Radial Measure: This measure (a.k.a Debreu- Farrel measure ofCCR / BCC measure) indicates the necessary improvements when all relevant factors are improved by the same factor equiproportionally.

Benchmarks : (a) For inefficient DMU the reference DMUs with corresponding intensities.
(b) For efficient DMU the number of inefficient DMU which have chosen the DMU as benchmark.

Table 4.1 Container Terminal Efficiency ratings

Table 4.1 Container Terminal Efficiency ratings

	DUM	Score	Benchmarks
1	SANTA CRUZ DE CT	95.77%	34 (0.96)
2	TCP CT (BRAZIL)	80.58%	34 (0.25) 42 (0.09) 69 (0.46)
3	FOS CT (FRANCE)	29.81%	39 (0.00) 41 (0.00) 42 (0.06) 45 (0.00) 54 (0.02) 60 (0.21)
4	EURO GATE	88.79%	8 (0.10) 34 (0.69) 59 (0.03)
5	BURCHARDKAI	37.35%	8 (0.06) 29 (0.10) 59 (0.21)
6	EURO GATE	11.00%	8 (0.03) 19 (0.08) 42 (0.00) 59 (0.02) 67 (0.02)
7	VENIZELOS CT	86.59%	8 (0.06) 19 (0.63) 29 (0.10) 34 (0.34) 42 (0.06)
8	VOLTRI CT	100.00%	5
9	LA SPEZIA CT	62.08%	34 (0.06) 42 (0.40) 63 (0.16)
10	MARSAXLOKK CT	67.28%	45 (0.01) 63 (0.62)
11	APM TERMINAL	61.01%	39 (0.06) 54 (0.13) 63 (0.42)
12	HT HOLLAND CT	72.98%	41 (0.15) 42 (0.37) 53 (0.14) 59 (0.03)
13	SANTA APOLONIA CT	74.55%	34 (0.19) 69 (0.19)
14	VICS	29.02%	34 (0.12) 39 (0.01) 42 (0.04) 63 (0.02)
15	TCB TERMINAL	40.68%	41 (0.01) 54 (0.01) 67 (0.10)
16	VALEN CINA PUBLICT	83.43%	8 (0.08) 29 (0.68) 34 (0.07)
17	TRINITY	57.92%	41 (0.16) 42 (0.22) 54 (0.10) 63 (0.10)
18	SOUTHAMPTON CT (UK)	68.49%	39 (0.07) 41 (0.61)
19	TESPORT CT(UK)	100.00%	2
20	SOUTHEND CT, UK	100.00%	4
21	VANTERN CT	16.92%	20 (0.07) 34 (0.00) 42 (0.08) 69 (0.02)
22	VIETNAM INT CT	34.86%	34 (0.09) 41 (0.01) 42 (0.09) 60 (0.09) 63 (0.07)
23	LONG BEACH CT	50.30%	34 (0.05) 42 (0.05) 63 (0.13)
24	YUSEN CT (USA)	32.26%	34 (0.19) 41 (0.01) 42 (0.02) 63 (0.09)
25	SANTO THOMAS CT	44.82%	42 (0.41) 54 (0.04)
26	VERACRUZ CT	71.41%	34 (0.17) 59 (0.10) 69 (0.11)
27	MANZA NILLO INT CT	65.92%	34 (0.22) 41 (0.13)
28	FREE PORT CT	62.68%	34 (0.12) 39 (0.17) 42 (0.16) 60 (0.12) 63 (0.06)
29	TCH TERMINAL	100.00%	7
30	KING STON CT (JAMAICA)	43.58%	29 (0.02) 36 (0.00) 42 (0.05) 59 (0.04) 67 (0.03)
31	POINT LISAS	62.41%	36 (0.45) 59 (0.18)
32	BEUNOS AIRES CT	92.59%	20 (0.31) 69 (0.00)
33	ITAJAI CT	14.98%	20 (0.04) 34 (0.01) 42 (0.05) 63 (0.03)
34	SAN ANTONIO CT	100.00%	27

	DUM	Score	Benchmarks
35	SWASON WEST CT	97.40%	34 (0.10) 42 (0.11) 69 (0.70)
36	(AUSTRALIA)	100.00%	3
37	FERGUESSON CT	60.59%	29 (0.30) 34 (0.04) 59 (0.01)
38	DALIAN CT (PRC)	84.54%	42 (0.65) 54 (0.19)
39	KWAI CHUNG(HK)	100.00%	7
40	BELIUN CT (PRC)	90.96%	34 (0.43) 41 (0.26) 63 (0.15)
41	WAIGAOQIAO CT (PRC)	100.00%	16
42	SHEKOU CT (PRC)	100.00%	29
43	CHENNAI CT (INDIA)	56.25%	34 (0.10) 42 (0.03) 63 (0.30)
44	JNP CT (INDIA)	84.36%	41 (0.11) 45 (0.11) 60 (0.43) 63 (0.19)
45	NHAVA SHEVA CT (INDIA)	100.00%	6
46	NCB CT (JAPAN)	100.00%	0
47	OMNI R1-5 CT (JAPAN)	95.21%	34 (0.01) 42 (0.11) 63 (0.14)
48	TAKA SAGO CT (JAPAN)	85.18%	36 (0.19)
49	SHIMIZU (JAPAN)	38.43%	20 (0.01) 57 (0.00) 63 (0.25)
50	MC 1,2TERMINAL (JAPAN)	48.63%	39 (0.02) 41 (0.03) 42 (0.13) 45 (0.10) 60 (0.04)
51	U-AM CT (S.KOREA)	80.54%	42 (0.03) 54 (0.03) 60 (0.74) 63 (0.01)
52	SHINSUNDAE CT (S.KOREA)	82.26%	34 (0.17) 41 (0.35) 42 (0.05) 63 (0.26)
53	KLANG CT 1&3 (MALAYSIA)	100.00%	1
54	TANJUNPELEPAS (MALAYSIA)	100.00%	9
55	QUASIM INT CT (PAKISTAN)	39.40%	34 (0.01) 41 (0.00) 63 (0.21)
56	MANILA INT CT	54.21%	41 (0.13) 42 (0.18) 54 (0.01) 63 (0.22)
57	JURONG CT (SINGAPORE)	100.00%	1
58	GT CT	71.77%	45 (0.29) 54 (0.08)
59	KELUNG CT	100.00%	10
60	ESCO CT	100.00%	7
61	LCIT CT	95.44%	41 (0.23) 45 (0.03)
62	ASHOD CT (ISRAEL)	49.94%	34 (0.14) 42 (0.20) 63 (0.10)
63	KHOR FAKKAN	100.00%	21
64	ADEN CT	12.63%	41 (0.01) 42 (0.04) 63 (0.02)
65	DAMIETTA CT	87.57%	34 (0.36) 42 (0.52)
66	CAPE TOWN CT (S.AFRICA)	84.62%	29 (0.16) 42 (0.03) 59 (0.04) 67 (0.20)
67	DURBAN CT (S.AFRICA)	100.00%	4
68	TANZA NIA CT	30.12%	34 (0.09) 39 (0.00) 42 (0.04) 60 (0.17)
69	MAURITIUS CT	100.00%	6
70	AQUABA CT	48.50%	29 (0.04) 34 (0.24) 59 (0.06)

4.1.1 Segmenting the Container Terminals by DEA Recursive Analysis

At first the overall efficiency score of 70 container terminals is measured upon which the efficient container terminal, with a score of 1, is placed in one group labeled as 'tier 1'.

Table 4.2 Segmented Container Terminals Tier 1

Efficient Tier 1	Score	Benchmark
VENIZELOS	100.00%	No benchmark
SOUTHAMPTON CT	100.00%	
TESPORT CT	100.00%	
FREE PORT CT	100.00%	
ITAJAI CT	100.00%	
SWASON WEST CT	100.00%	
KWAI CHUNG	100.00%	
WAIGAOQIAO CT	100.00%	
SHEKOU CT	100.00%	
NHAVA SHEVA CT	100.00%	
NCB CT	100.00%	
KLANG CT 1&3	100.00%	
TANJUN PELEPAS	100.00%	
JURONG CT	100.00%	
KELUNG CT	100.00%	
ESCO CT	100.00%	
KWOR FAKKAN CT	100.00%	
DURBAN CT	100.00%	
MAURI TIUS CT	100.00%	

After the first analysis, the efficient terminals with score one which has been grouped is excluded in the second stage of efficiency measurement. The second measurement produces a group of efficient terminals with a score of 1 and they are segmented with a label 'tier 2'.The same procedure is repeated until the number ofremaining terminals are at least three times greater than that of total inputs and outputs.

Table 4.3 Segmented Container Terminals Tier 2

Efficient Tier 1	Score	Benchmark
AARHUS CT	100.00%	34 (0.96)
FOS CT, MARSEILLES	100.00%	8 (0.10) 34 (0.69) 59 (0.03)
EURO GATE	100.00%	8 (0.06) 29 (0.10) 59 (0.21)
LA SPEZIA	100.00%	45 (0.01) 63 (0.62)
MARSAXLOKK	100.00%	39 (0.06) 54 (0.13) 63 (0.42)
APM TERMINAL	100.00%	41 (0.15) 42 (0.37) 53 (0.14) 59 (0.03)
HT HOLLAND CT	100.00%	34 (0.19) 69 (0.19)
VICS	100.00%	41 (0.01) 54 (0.01) 67 (0.10)
TCB TERMINAL	100.00%	8 (0.08) 29 (0.68) 34 (0.07)
TRINITY	100.00%	39 (0.07) 41 (0.61)
SANTOTHOMAS CT	100.00%	34 (0.17) 59 (0.10) 69 (0.11)
KINGSTON CT	100.00%	36 (0.45) 59 (0.18)
POINT LISAS	100.00%	20 (0.31) 69 (0.00)
SAN ANTONIO CT	100.00%	34 (0.10) 42 (0.11) 69 (0.70)
DALIAN CT	100.00%	42 (0.65) 54 (0.19)
BELIUN CT	100.00%	34 (0.43) 41 (0.26) 63 (0.15)
JNP CT	100.00%	41 (0.11) 45 (0.11) 60 (0.43) 63 (0.19)
OMNI R1-5 CT	100.00%	34 (0.01) 42 (0.11) 63 (0.14)
TAKA SAGO CT	100.00%	36 (0.19)
U-AM CT	100.00%	42 (0.03) 54 (0.03) 60 (0.74) 63 (0.01)
SHIN SUNDAE CT	100.00%	34 (0.17) 41 (0.35) 42 (0.05) 63 (0.26)
GT CT	100.00%	45 (0.29) 54 (0.08)
LCIT CT	100.00%	41 (0.23) 45 (0.03)
DAMIETTA CT	100.00%	34 (0.36) 42 (0.52)
CAPE TOWN CT	100.00%	29 (0.16) 42 (0.03) 59 (0.04) 67 (0.20)
AQUABA CT	100.00%	29 (0.04) 34 (0.24) 59 (0.06)

After the third analysis the results are summarized in table 4.4 with their respective Benchmarks which are filtered from the overall efficiency to prevent biased reference numbers due to the elimination of efficient terminals in the respective tiers. The fourth tier produced by the recursive analysis of DEA concludes the segmentation of container terminals based on their efficiency levels. Table 4.5 indicates the last tier 4 with their benchmarks.

Table 4.4 Segmented Container Terminals Tier 3

Efficient Tier 1	Score	Benchmark
SANTA CRUZ DE CT	100.00%	1 (0.08) 35 (0.80) 65 (0.02)
EURO GATE	100.00%	18 (0.05) 40 (0.28) 45 (0.03)
VOLTRI	100.00%	1 (0.12) 35 (0.08) 51 (0.58)
VALENCINA PUBLIC CT	100.00%	18 (0.28) 40 (0.35) 47 (0.34) 52 (0.06)
VIETNAM INT CT	100.00%	32 (0.34) 35 (0.02) 40 (0.14) 47 (0.19)
VERACRUZ CT	100.00%	11 (0.41) 47 (1.11) 52 (0.10)
MANZANILLO INT CT	100.00%	11 (0.41) 47 (1.11) 52 (0.10)
TCH TERMINAL	100.00%	16 (0.08) 48 (0.21) 51 (0.17) 66 (0.04)
FERGUESSON CT	100.00%	16 (0.40)
CHENNAI CT	100.00%	32 (0.15) 47 (1.37)
ASHOD CT	100.00%	10 (0.01) 32 (0.15) 47 (0.70) 65 (0.24)

Table 4.5 Segmented Container Terminals Tier 4

Efficient Tier 1	Score	Benchmark
TCP CT BRAZIL	100.00%	17 (0.21)
BURCH ARDKAI	100.00%	5 (0.22) 9 (0.00) 17 (0.04) 37 (0.08)
SANTA APOLONIA CT	100.00%	17 (0.05) 28 (0.09) 62 (0.05)
SOUTH END CT	100.00%	2 (0.02) 23 (0.14) 27 (0.16) 62 (0.03)
VANTERN CT	100.00%	17 (0.16) 28 (0.05) 62 (0.18)
LONG BEACH CT	100.00%	2 (0.23) 9 (0.05) 17 (0.13)
YUSEN CT	100.00%	2 (0.31) 17 (0.42) 62 (0.10)
BEUNOS AIRES CT	100.00%	23 (0.17) 27 (0.08) 28 (0.02) 62 (0.04)
SHIMIZU	100.00%	43 (0.35) 62 (0.36)
MC 1,2TERMINAL	100.00%	17 (0.23) 28 (0.18)
QUASIM INT CT	100.00%	23 (0.68) 27 (0.07) 28 (0.06) 62 (0.02)
MANILA INT CT	100.00%	17 (0.75) 28 (0.00) 62 (0.09)
ADEN CT	100.00%	17 (0.09)
TANZANIA CT	100.00%	17 (0.05) 28 (0.11)

4.1.2 Generating classification Rules for Each Tier

A classification rule using See 5 prepares a training set of cases, each described in terms of the given attributes (seven inputs) and a known class i.e.tier number. These cases come from a source such as container terminal tiers as a result of recursive analysis of DEA. The induction process of See 5 attempts to find a method of classifying a case, expressed as a function of the attributes that explains the training cases and that may also be used to classify unseen cases.

There are four classes (1,2,3,4) that have been identified by the recursive analysis and the seven factors input along with the derived classes that influence the class or decision in this experiment and they are as follows:

+ Quay length (QL)
+ Terminal area (TA)
+ Quay cranes (QC)
+ Transfer cranes (TC)
+ Straddle carriers (STC)
+ Reach stackers (RSC)
+Labor hours (LBH)
+ Tiers

The values of the above factors are shown in table 4.6. The arrangement of the values dictates from left to right with eight factor values and tier numbers that are arranged in rows. The 70 container terminals are utilized to train the discriminant descriptor See 5.

Table 4.6 Training cases for See 5

	QL{I}	TA{I}	QC{I}	TC{I}	STC{I}	RSC{I}	LBH{I}	Tier
AARHUS CT	1000	530000	10	0	0	36	24	2
SANTA CRUZDE	900	481000	7	2	0	11	24	3
TCP CT BRAZIL	655	292300	5	14	0	2	24	4
FOS CT	1180	330000	4	0	8	27	24	2
EURO GATE	3946	1450000	19	0	66	8	24	2
BURCHARDKI	2850	1600000	18	4	94	7	24	4
CT 70*	···	······	······	······	······	·····	·····	······

The output of the decision tree generator for the cases of container terminals is shown in figure 4.1. It is to be noted that the numbers at the leaves, of the form (N) or (N/E), Nis the sum of the fractional cases that reach the leaf; E is the number of cases that belongs to classes other than the nominated class.

Decision tree:

```
TC <= 3: abovetier4 (25)
TC > 3:
:...TC > 28: abovetier4 (11)
    TC <= 28:
    :...QL <= 520: abovetier4 (5)
        QL > 520:
        :...QC <= 7:
```

```
:...TC <= 15: tier4 (18)
:   TC > 15:
:   :...QL <= 770: tier4 (2)
:       QL > 770: abovetier4 (2)
QC > 7:
:...LBH <= 22: abovetier4 (2)
    LBH > 22:
    :...STC > 73: tier4 (2)
        STC <= 73:
        :...STC > 13: abovetier4 (2)
            STC <= 13:
            :...QC <= 8: abovetier4 (3)
                QC > 8:
                :...QC > 11: abovetier4 (3)
                    QC <= 11:
                    :...TA <= 310000: abovetier4 (2)
                        TA > 310000: tier4 (6)
```

Figure 4.1 Induced Decision Tree by See 5for discriminating Tier 3 and Tier 4

Decision trees are usually simplified by discarding one or more sub- trees and replacing them with leaves; while building trees, the class associatedwith a leaf is found by examining the training cases covered by leaf and choosing the most frequent class. See 5 also allow replacement of a sub-tree by one of the branches. Figure 4.2 shows a decision tree after pruning operation. Pruning a decision tree might cause misclassification of the training cases but it is done by producing more comprehensible tree structures and finally simpler production rules without compromising accuracy on classifying unseen cases and after the production of decision tree, classification rule can be extracted.

In addition to generating classification rules, See 5 can discover major input and output variables affecting the efficiency of the units. It can also find the order of influences of the respective values, e.g. the sequence such as RSC, LBH, TA , STC,TC, QC, QL as shown in figure 4.2 and it can be inferred that the effect of RSC (Reach Stackers) to the efficiency of container terminals is greater than that of Labor Hours (LBH).

A simplified decision tree with sequences is shown in figure 4.2.

Decision tree:

```
RSC <= 0:
:...LBH <= 20: belowtier2 (2)
:   LBH > 20: abovetier2 (14)
RSC > 0:
:...TA > 926100: belowtier2 (9)
```

```
TA <= 926100:
:...STC > 53: abovetier2 (7)
    STC <= 53:
    :...LBH <= 22: belowtier2 (8)
        LBH > 22:
        :...TC > 64: abovetier2 (3)
            TC <= 64:
            :...TA <= 130000: abovetier2 (8/1)
                TA > 130000:
                :...RSC > 10: belowtier2 (13)
                    RSC <= 10:
                    :...TA <= 255600: belowtier2 (8)
                        TA > 255600:
                        :...TA > 500000: belowtier2 (4)
                            TA <= 500000:
                            :...TA > 450000: abovetier2 (5)
                                TA <= 450000:
                                :...TA > 300000: belowtier2 (6)
                                    TA <= 300000:
                                    :...QC > 7: abovetier2 (6)
                                        QC <= 7:
                                        :...QL <= 600: abovetier2 (3)
                                            QL > 600: belowtier2 (2)
```

Figure 4.2 Simplified Decision Tree for discriminating Tier 1 and Tier 2

4.2 Determining the Improvement Path of Inefficient Terminals

4.2.1 Clustering the Terminals using SOM

Clustering the terminals using SOM is divided into two steps. The first step is to train the SOM against the terminals as a training data set. The second one is to map input DMUs to output DMU clusters.

1) To train the SOM

Training algorithms of a SOM adjusts the weights and thresholds using a set of training patterns. The iterative gradient-descent training algorithms, including the SOM, attempt to reduce the training error on each epoch. It specifies different stopping conditions as to when training should stop. The simplest condition is that training should stop after a set number of epochs, or iteration. This is the most commonly used condition.

2) To map input terminals to output terminal clusters

SOM is designed for unsupervised clustering of data; i.e. it is given training pattern which only contain inputs, the SOM assigns output units which represent cluster to inputs. Once a

SOM has been trained, it forms a topological map using the output layer. The mapping inputs the terminal patterns to the output terminal cluster reflecting the existing similarities in the inputs. The self-organizing algorithm not only assigns cluster centers to terminals, but also tends to group similar centers on terminals close to each other. Figure 4.3 shows the result of clustering 70 container terminals where four clusters are formed as a result of output. The numbers in each cluster indicates respective container terminals.

Cluster (1) 34, 50, 51, 52, 53, 60, 63, 68	**Cluster (2)** 1, 2, 3, 4, 5, 6, 10, 12, 15, 21, 31, 44, 64
Cluster (3) 7, 8, 11, 13, 14, 16, 17, 18, 19, 22, 23, 24, 26, 27, 28, 29, 32, 33, 35, 36, 37, 38, 39, 40, 42, 43, 45, 47, 49, 56, 57, 58, 62, 65, 67, 69, 70	**Cluster (4)** 9, 20, 41, 46, 48, 54, 55, 59, 61

Figure 4.3 Result of SOM Analysis for Container Terminals

Table 4.7 and 4.8 summarizes the characteristics of each cluster in details. The Table 4.8 shows four clusters and in them the number of container terminals which denotes the size of the cluster their position on grid and average value of input and output.

Table 4.7 SOM Clustering of Container Terminals

Clustering using Self Organizing Map	
Number of variables used for clustering	7
Number of observations used for clustering	70
Number of Clusters	4

Table 4.8 Cluster Sizes, Position and Means of Container Terminals

Cluster Sizes					
		Cluster 1	Cluster 2	Cluster 3	Cluster 4
		9	13	39	9
Cluster Position on the grid					
		Cluster 1	Cluster 2	Cluster 3	Cluster 4
	Row	1	1	2	2
	Column	1	2	1	2
Cluster means					
	Overall	Cluster 1	Cluster 2	Cluster 3	Cluster 4
QL {I}	1322.8	715.3	1309.3	1156.1	978.3
TA {I}	847530.9	322918.2	607207.7	510986.4	613655.6
QC {I}	9.7	6.1	9.6	8.9	11.4
TC {I}	14.2	15.2	8.3	15.1	17.9
STC {I}	11.0	8.0	20.8	12.2	10.0
RSC {I}	11.6	9.2	9.8	8.7	4.9
LBH {I}	22.5	23.1	23.3	22.5	22.2

4.2.2 Determining the Benchmark of inefficient Terminals

The DEA recursive analysis produces four segments of terminals based on their efficiency level. SOM on the other hand clusters the container terminals based on their input traits. After organizing the terminals based upon these procedures the projection of inefficient terminal can be determined. The inefficient terminals in the lowest tier have benchmark on its immediate upper tier due to similar features grouped by implementing SOM. Similar is the case with the terminals in tier 3 or 2 belonging to separate clusters.

4.2.3 Improvement Projection

After obtaining all the results upon applying the tools DEA, SOMand Decision Tree to 70 container terminals we can finally analyze their efficiency and inefficiency level and after clustering them the improvement projection is decided. The decision tree helps to indicate the variables at each efficiency level that is significant for improvement in order to get promoted to its upper tier. The analysis after application reveals the improvement projection of container terminal for example we take the Shimizu container terminal which is in the lowest tier that is tier 4 and its reference terminal in tier 3 is Ashod container terminal. For Ashod container terminalreference set is CapeTtown container terminal in tier 2 and subsequently in tier 1 is Kwai Chung container terminal as a benchmark for Cape Town container terminal. The benchmarking of the above terminals is the resultant of cluster 3, obtained from SOM analysis and the same procedure is applicable to other three clusters. The decision tree reveals the significance of each variable at the respective tier. The results indicate that for Shimizu container terminal, TC (Transfer Cranes) is the most significant value followed by QL (Quay Length) for its improvement to get promoted to tier 3. Similarly, for terminals in tier 3 is RSC (Reach Stackers) followed by TA (Terminal Area) and QC (Quay Cranes) and in tier 2 the most significant input is RSC (Reach Stackers) again followed by LBH (Labor Hours).

5. Conclusion

In this paper DEA is used to calculate the efficiency of container terminalsand at the same time implemented for segmenting the terminals after estimation of overall relative efficiency. DEA as a multi-factor productivity measurement model is used to measure efficiency and set benchmarks for the inefficient terminals. But the benchmark that is derivedby linear combination of units which utilizes less input than the test DMU while maintaining at least the same output level may not be inherently similar. To overcome this problem, in this paper, two fields of information technology DEA and data mining is integrated to achieve a synergy-producing result that cannot be obtained if each model is to operate individually.

The research design proposed here is to set proper benchmark and improvement projection for the terminals which otherwise is not possible with DEA alone. Its application to container port industry, gives a valuable insight regarding the improvement projection for the inefficient terminals in terms of its technical inputs.Since container terminals are heterogeneous and there is a need to evaluate these units under the DEA due to its acceptance as a performance measurement, this paper modifies the DEA to work with these heterogeneous units. The units involved in the construction of composite DMU can be utilized as benchmarks for improving the inefficient DMU. DEA also allows for computing the necessary improvements required in the inefficient unit's inputs and outputs to make it efficient. Although benchmarking in DEA allows for the identification of targets for improvements, it has certain limitations. An inefficient DMU and its benchmarks may not be inherently similar in their operating practices. Hence, the fusion of unsupervised learning tool Kohonen's SOM and decision tree analyzer See 5 showed some valuable results in relation to benchmarking of the container terminals. The results throw light on the need to upgrade the DEA tool used for benchmarking and improvement projection which can be achieved by the proposed methodology.

Inefficiency and benchmark of a container terminal cannot be evaluated alone by technical inputs. Some 'environmental factors' also affects the efficiency and future research can include such variables for efficiency measure which can throw some light on real reasons behind port inefficiency. Cross-sectional data used in this research may produce some misleading result upon which panel data is recommended which can provide a comprehensive picture on port efficiency. However, further research focusing on data quality, fusion techniques and specific characteristics of container terminals might give interesting insight.

References

1. Banker R.D. (1993), "Maximum Likelihood, Consistency and Data Envelopment Analysis: A statistical Foundation," Management Science, Vol. 39, No. 10 pp. 1265-1272.
2. Charnes, A., Cooper, W.W., and Rhodes, E. (1978), "Measuring the Efficiency of Decision Making Units," European Journal of Operational Research 2, 429-444.
3. Coto,P, Banos, J, and Rodriguez, A., 2000: "Economic Efficiency in Spanish Ports. Some empirical evidence," Maritime Policy and Management, 27, pp.169-174.
4. Dula, J. H., N. Venugopal (1995),"On Characterizing the Production Possibility Set for the CCR Ratio Model in DEA," Int. J. System SCI., Vol. 26, No. 12, pp.2319-2325.
5. E. Thanassoulis (1996), A Data Envelopment Analysis Approach to Clustering Operating Units for Resource Allocation Purposes" Omega, Int. J. Mgmt. Sci, 24/4, 463-467.
6. Farrel, M.J. (1957), "The Measurement of Productive Efficiency," Journal of the Royal Statistical Society 120, Series A, Part III, 253-281.
7. Herrero,I., Sean Pascoe, (2002) "Estimation of Technical Efficiency: a review of some of the stochastic frontier and DEA software," Economics Network: Vol. 15, Issue 1.
8. Herrero,I., Jose. L.S., (2005) "Using the DEA methodology to rank software technical efficiency," Communications of the ACM: Vol. 48. Issue1. pp 101-105.
9. Jaffar.W.D, Gordo Berry, Ian Ridley (2005) "Improving Performance in Port Authorities," Proceedings of Integrating for Excellence Conference, June, UK.
10. Kevin Cullinane, Dong-Wook Song, Ping Ji. Wang.T.F.(2004), "An Application of DEA Windows Analysis to Container Port Production Efficiency", Review of Networks Economics Vol.3,184-206.
11. Kevin Cullinane, Dong-Wook Song, Wang.T.F, (2003) "Container Port Production Efficiency: A comparative study of DEA and FDH approaches,"Journal of the Eastern Asia Society for Transportation Studies Vol. 5, pp 698-793.
12. Kevin Cullinane, Ping Ji. Wang.T.F, (2005) "The relationship between privatization and DEA estimates of efficiency in the container port industry," Journal of Economics and Business: pp1-30.
13. Liu, Z. 1995. The comparative performance of public and private enterprises: The case of British ports. Journal of Transport Economics and Policy, 29(3), 263-274
14. Marlow,P. and Paixao, A.C. (2002) "Measuring Lean Ports Performance," Proceedings of International Association of Maritime Economists Conference, November, Panama, 13-15.
15. Martinez-Budira, E., Diaz-Armas,R., Navarro-Ibanez, M. and Ravelo-Mesa, T. (1999) "A Study of the Efficiency of Spanish Port Authorities Using Data Envelopment Analysis,"International Journal of Transport Economics, XXVI: 37-253.
16. Roll and Hayuth Y. (1993) "Port Performance Comparison Applying Data Envelopment Analysis (DEA)," Maritime Policy and Management, 20: 153-161.
17. Seiford, L.M., and Thrall, R.M.(1990), "Recent Development in DEA: The mathematical programming approach to frontier analysis", Journal of Econometrics 46, 7-38.

18. Sengupta, J. K., (1997), "Contributions to Data Envelopment Analysis," Cybernetics and Systems: An International Journal, Vol. 28, pp. 79-97.

19. Serranco Cinca; C (1998): "Self-organizing Maps for Initial Data Analysis: Let Financial Data Speak for themselves," in visual intelligence in Finance using Self-organizing Maps, July 1998, Ed Guido Deboech & Teuvo Kohonen, Chapter 7.

20. Sturart Russel, Peter Norvig, (2002) "Artificial Intelligence: A modern Approach" Section 18, 3; page531.

21. Talley, W.K. (1998) "Optimum Throughput and Performance Evaluation of Marine Terminals," Maritime Policy and Management, 15: 327-331.

22. Talluri, S., J.Sarkis (1997), Extensions in Efficiency Measurement of Alternate Machine Component Grouping Solution via Data Envelopment Analysis,"IEEE Transactions on Engineering Management, Vol. 44, No. 3, pp. 299-304.

23. Tongzon, J. L. (1995) "Determinants of Port Performance and Efficiency," Transport Research A, 29: 245-352.

24. Tongzon, J.L. (2001) "Efficiency Measurement of Selected Australian and Other International Ports Using Data Envelopment Analysis," Transportation Research A: Policy and Practice 35: 113-128.