Received: 6 June 2019 | Revised: 11 September 2019 | Accepted: 9 October 2019

DOI: 10.1111/coin.12251

SPECIAL ISSUE ARTICLE

Computational Intelligence WILEY

Efficiency analysis for stochastic dynamic facility layout problem using meta-heuristic, data envelopment analysis and machine learning

Akash Tayal[1] | **Utku Kose**[2] | **Arun Solanki**[3] | **Anand Nayyar**[4] | **José Antonio Marmolejo Saucedo**[5]

[1]Department of Electronic and Communication, Indira Gandhi Delhi Technical University for Women, Delhi, India

[2]Department of Computer Engineering, Suleyman Demirel University, Isparta, Turkey

[3]Department of CSE, School of ICT, Gautam Buddha University, Greater Noida, India

[4]Graduate School, Duy Tan University, Da Nang, Vietnam

[5]Faculty of Engineering, Universidad Panamericana, CDMX, Mexico

Correspondence
Utku Kose, Department of Computer Engineering, Suleyman Demirel University, Isparta, Turkey.
Email: utkukose@sdu.edu.tr

Abstract

The facility layout problem (FLP) is a combinatorial optimization problem. The performance of the layout design is significantly impacted by diverse, multiple factors. The use of algorithmic or procedural design methodology in ranking and identification of efficient layout is ineffective. In this context, this study proposes a three-stage methodology where data envelopment analysis (DEA) is augmented with unsupervised and supervised machine learning (ML). In stage 1, unsupervised ML is used for the clustering of the criteria in which the layouts need to be evaluated using homogeneity. Layouts are generated using simulated annealing, chaotic simulated annealing, and hybrid firefly algorithm/chaotic simulated annealing meta-heuristics. In stage 2, the nonparametric DEA approach is used to identify efficient and inefficient layouts. Finally, supervised ML utilizes the performance frontiers from DEA (efficiency scores) to generate a trained model for getting the unique rankings and predicted efficiency scores of layouts. The proposed methodology overcomes the limitations associated with large datasets that contain many inputs / outputs from the conventional DEA and improves the prediction accuracy of layouts. A Gaussian distribution product demand dataset for time period $T = 5$ and facility size $N = 12$ is used to prove the effectiveness of the methodology.

KEYWORDS

data envelopment analysis, intelligent optimization, machine learning, stochastic dynamic facility layout problem

1 | INTRODUCTION

The facility layout problem (FLP) is defined as the arrangement of facilities "N" in a given space "M" to get an optimal layout so that the material travel distance is minimized. Material movement reduction lowers work-in-process levels and throughput times, lessens the damage to the product, streamlines material control and scheduling, and eases the congestion in the area thus minimizing material handling costs (MHC). Tompkins et al[1] discussed that MHC contributes between 20% and 50% of the product's operating cost so that it can be used to evaluate the effectiveness of a layout. An efficient layout minimizes the distance traveled by the material between different locations which, in-turn, reduces the MHC. MHC is often referred to as the quantitative factor that impacts the facility's layout. FLP details can be found in Rosenblatt and Lee,[2] Kusiak and Heragu,[3] and Meller and Gau.[4] Koopmans and Beckman[5] represented FLP using the quadratic assignment problem (QAP), which is a combinatorial optimization problem (COP). In QAP modeling, FLP is a discrete representation where the equal-sized facilities are assigned to the same number of known locations. In practical scenarios, facility layout design is dependent on the demand that can be static, dynamic, or random and time horizon that can be single period or multiple periods. The paper discusses uncertain product demand over multiple time periods, known as stochastic dynamic FLP (SDFLP), and its solution methodology. The uncertain product demand due to social, economic, political, and governmental rules and regulations and environmental and seasonal conditions is unavoidable. This results in various qualitative factors, such as the working conditions, storage, waste management, safety regulations, and ease of maintenance, which impact the facility's layout. Due to the presence of quantitative parameters and qualitative factors, traditional methods, such as the weight aggregation method, Pareto method, multi-attribute decision-making and algorithmic approaches (meta-heuristics) to solve SDFLP, may not be suitable. As a result, there is a need to develop a solution methodology that considers the characteristics of SDFLP such as the problem size, linearity, and congruency of various criteria.

The performance of any system is measured by its efficiency. Data envelopment analysis (DEA) is a generally used method for assessing the relative efficiencies of decision-making units (DMUs). DEA has some advantages in that it does not require any suppositions to be made about the distribution of inefficiency or a particular functional form of the data to determine the efficient DMUs or efficient frontiers. At the same time, DEA suffers from some limitations, such as the dimensionality curse, its inability to efficiently rank unique DMUs and its inability to efficiently predict new DMUs without recalculation. Machine learning (ML) is a fairly new domain that has gained a lot of importance as it gives computers the ability to learn and make accurate predictions without being explicitly programmed. ML architecture builds an algorithm that receives input data and uses various statistical techniques to predict output within the acceptable limits. ML centers around the recognition of patterns, regularities in data as well as systematic relationships between the variables.

The evolution of technology facilitates and encourages us to find new ways to solve SDFLPs. One such approach is the combination of two or more methods (procedural and/or algorithmic) to overcome the shortcomings of a single method. Thus, combining DEA with ML can strengthen the capabilities of DEA to rank and predict the efficiencies of the DMUs. The objective of this study is to use meta-heuristic, DEA and ML-based integrated methodology to solve SDFLP and identify the most efficient layout. The methodology proposed here consists of a three-stage process where DEA is augmented with unsupervised and supervised ML. In detail, an unsupervised ML is run for the clustering of the criteria in which the layouts need to be evaluated using homogeneity. Next,

TABLE 1 Acronyms used in the paper

Acronym	Description
FLP	Facility layout problem
MHC	Material handling costs
QAP	Quadratic assignment problem
COP	Combinatorial optimization problem
SDFLP	Stochastic dynamic facility layout problem
MADM	Multi attribute decision-making
DEA	Data envelopment analysis
DMUs	Decision-making units
ML	Machine learning
ANN	Artificial neural network
COFAD-F	Computerized facility design flexible heuristic
SA	Simulated annealing
AHP	Analytical hierarchal process
NN	Neural networks
PNNs	Probabilistic neural networks
BPNN	Back-propagation neural network
CCR-CV	Charnes, Cooper, and Rhodes—cluster validity
PDF	Probability distribution function
TS	Tabu search
CSA	Chaotic simulated annealing
GA	Genetic algorithm
ACO	Ant colony optimization
PSO	Particle swarm optimization
FA	Firefly algorithm
CCR	Charnes, Cooper, and Rhodes
BCC	Banker, Charnes, and Cooper

the layouts are generated using simulated annealing (SA), chaotic SA (CSA), and hybrid firefly algorithm/chaotic simulated annealing meta-heuristics. Here, the nonparametric DEA approach is also used to identify efficient and inefficient layouts and a supervised ML is then employed for the performance frontiers from DEA (efficiency scores) to generate a trained model that is used to get the unique rankings and predicted efficiency scores of layouts. In the study, the background regarding the developed methodology, its technical details, and evaluation findings in the context of the SDFLP have been provided generally.

Moving from the subject of the study and the research, the remaining content is organized as follows: Background information regarding a review of the literature and the motivation of the study is provided in Section 2. Following to that, the Section 3 explains the SDFLP mathematical formulation, meta-heuristics, DEA, and ML. Next, the Section 4 describes the proposed meta-heuristic-DEA-ML integrated methodology for solving SDFLP, along with the purpose and function of each step. Section 5 gives the numerical illustration for dimension reduction, training, validation, ranking, and prediction of SDFLP. Finally, the Section 6 concludes the research findings and gives the future scope of the framework. The acronyms used within the paper are indicated in Table 1.

2 | BACKGROUND

In this section, the past literature on facility layout with an emphasis on SDFLP is given first. Furthermore, a past overview on DEA, artificial neural network (ANN), and ML-oriented view are given accordingly, by also discussing about motivation of this study. In this way, it was aimed to inform the readers about background by opening minds about the associated literature and the origin of the study done here.

2.1 | Literature review

First, it is important to start from the origin of the problem considered in this study. Shore and Tompkins[6] presented the concept of a flexible layout in a stochastic environment. They developed computerized facility design flexible heuristic (COFAD-F). Rosenblatt and Lee.[2] Rosenblatt and Kropp.[7] and Palekar et al[8] solved the SDFLP with the objective to minimize MHC and rearrangement cost. Moslemipour and Lee[9] solved SDFLP using SA. Recent work on SDFLP can be studied from Tayal and Singh.[10-12] On the other hand, the initial models of the DEA technique were proposed by Charnes et al[13] and Banker et al.[14] They introduced DEA for assessing efficiency in a variety of organizations, including banks, hospitals, airlines, and universities. Modifications in DEA can be understood from the studies by Charnes et al,[15] and Tone.[16] Also, other DEA models can be seen in Emrouznejad et al.[17] Yang and Kuo[18] proposed an integrated multivariate and multiple attribute analysis approach based on the analytical hierarchal process and DEA for solving FLP. Azadeh et al[19] solved the flow shop FLP using fuzzy DEA. Tayal and Singh[20] presented an integrated approach of SA-DEA-technique for order preference by similarity to ideal solution (TOPSIS) for solving SDFLP. Tayal et al[21] used meta-heuristics along with DEA and proposed an integrated ranking approach for solving sustainable SDFLP.

As it can be understood, the literature has already been interested in using artificial intelligence for solving the related problem. Here, especially ML techniques have popularity in terms of alternative research processes. Athanassopoulos and Curram[22] first introduced the idea of the combination of neural networks (NNs) and DEA for classification and/or prediction. Other papers that presented the benefits of combining DEA and NN are Liu et al,[23] Liao et al,[24] Santin,[25] Santin et al[26] and Wang.[27] Azadeh et al[28] proposed a highly unique flexible ANN algorithm to measure and rank scores of Iranian steam power. Mostafa[29] combined probabilistic neural networks (PNNs) with DEA to predict the efficiency of banking systems. Wu[30] applied a DEA-Back-Propagation Neural Network (BPNN) to model efficiency scores of suppliers. Kwon[31] used ANN and DEA in application for the smart phone industry. Recent papers on the hybrid DEA-ANN by Hanafizadeh et al[32] used a NN back-propagation DEA for the measurement of mutual fund efficiency and showed a considerable reduction in computer memory and CPU time utilization as compared to conventional DEA methods. Misiunas et al[33] proposed deploying DEA to preprocess the data to remove outliers. This preserved monotonicity and reduced the size of the dataset used to train the ANN.[34] In this paper new cluster validity index, named Charnes, Cooper & Rhodes—cluster validity (CCR-CV), by integrating eight internal clustering efficiency measures based on DEA is presented. De Clercq et al[35] proposed an integrated framework of DEA and ML for accurate prediction of biogas production in an industrial-scale of Chinese unit.

2.2 | Motivation of the study

DEA is widely used in industry to benchmark and evaluate the efficiency. It continues to attract researchers because of its performance assessment ability. Although it is an attractive research topic, the application of DEA to the FLP is very limited. To identify an efficient facility layout remains an important requirement for decision make. As it was explained in the previous section and understood from the literature, there is a remarkable need to merge DEA with ML for statistical and regression analysis, achieving better results. Here, it is also important to discuss the problem from other factors. Briefly, the manufacturing units faces new regulations, industry standards and public policies infliction, new technologies introduction, and various other organizations improvement on a continuous basis. As a result, plant managers need to have the capability to respond to such challenges by utilizing sound performance measurement techniques. Efficiency evaluation and prediction of the facility layout is a useful and important data for the managers as it may affect their decision-making process. Hence, the academia is focusing their work and research in this direction to find different techniques and methodologies which may reduce the subjectivity in decision-making. One of the techniques is to use a stage-wise approach, where the problem is decomposed into number of stages (steps). This approach has many advantages such as: it improves the performance, it reduces the system complexity and it increases the accuracy of the results. Thus, the motivation of the study introduced here is to propose this new three-stage approach for solving SDFLP.

Moving from the mentioned motivation, a novel model called integrated metaheuristic-DEA-ML is proposed for measuring and predicting the efficiency of layouts based on the conflicting criteria of an SDFLP. As supporting the motivation, the purpose of the study is to present a comprehensive performance evaluation framework which is built on an intelligent decision support system. This frame work will help predict and identify an efficient layout based on managerial, social, and environmental criteria which can perform efficiently for a long term. Figure 1 depicts the meta-heuristic-DEA-ML integrated model as follows:

- In the first stage (**Criteria**), a large set of parameters/criteria influencing the facility layout design are identified. Then, unsupervised learning is used to remove redundant/correlated criteria and to obtain a linear combination of independent criteria without impacting the objective of SDFLP.

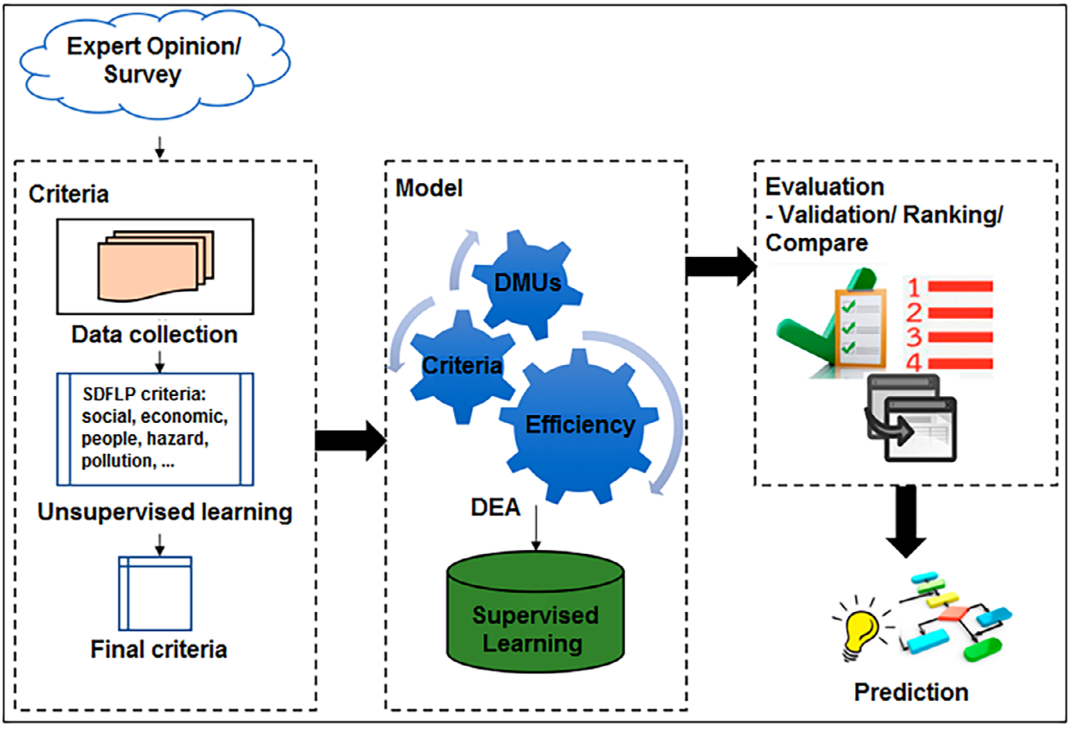

FIGURE 1 Metaheuristic-data envelopment analysis-machine learning integrated model for solving stochastic dynamic facility layout problem

- In the second stage (**Model**), the layout pool (DMUs) is generated using meta-heuristic, which is evaluated using the identified criteria to find the efficiency of each layout using DEA. This dataset is used by various ML algorithms for training and testing to get a final trained model that is used to predict the efficiencies of layouts.
- Finally, in the last stage (**Evaluation**), the ranking of layouts and comparison of results from various algorithms are performed. Also, the efficiency of new layouts (dataset) can be predicted in this stage.

3 | MATHEMATICAL FORMULATION AND THE METHODOLOGY

3.1 | Stochastic FLP

The product flow between facilities is an expression of demand that could be static, dynamic, or uncertain. Due to an uncertain product demand that can be modeled as stochastic random variables, the stochastic FLP (SFLP) has gained prominence. This random variable is expressed) with known mean and variance as a probability distribution function. The SDFLP mathematical model as discussed in Moslemipour and Lee[9] is given below, and the notations are described in Figure 2. Product demand with known mean and variance is assumed to be normally distributed. For deriving the mathematical model it is assumed that the distance between locations is numeric value, the facilities are equal size, and facilities are assigned discretely to each location.

Notations	Description
i, j	Indices for facilities $(i, j = 1,2,3\ldots N)$; $i \neq j$
l, q	Indices of locations $(l, q = 1,2,3\ldots N)$; $l \neq q$
f_{ij}	Flow of material between facilities, i to j
d_{lq}	Distance between locations, l to q
N	Number of facilities
$C(\pi)$	Total MHC for π-th layout
$E(\pi)$	Expected value of the π-th layout
$Var(\pi)$	Variance of the π-th layout
$Pr(\pi)$	Probability of the π-th layout
Z_p	Standard Z (random variable) value for percentile p
$U(\pi,p)$	Upper bound of $C(\pi)$with confidence level p
K	Index for parts $(k = 1, 2, \ldots, K)$
M_{ki}	Operation number for the operation done on part k by facility i
D_{kt}	Demand for part k in period t
B_k	Transfer batch size for part k
C_{tk}	Movements cost for part k in period t
Z	Random variable
a_{tilq}	Fixed cost of rearranging facility i from location l to location q in period t

FIGURE 2 Notations for stochastic dynamic facility layout problem mathematical model

In Equation (1), the material handling cost (MHC) is depicted by the first expression and the rearrangement cost (RA_c) is shown by the second expression.

$$\begin{aligned}\text{TMHC} = C_{iljq} = \text{Minimize}\Bigg\{&\left[\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{K}\frac{E(D_{tk})}{B_k}C_{tk}\sum_{l=1}^{N}\sum_{q=1}^{N}d_{lq}x_{til}x_{tjq}\right.\\ &\left.+ Z_p\sqrt{\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{K}\frac{\text{Var}(D_{tk})}{B_k^2}C_{tk}^2\left(\sum_{l=1}^{N}\sum_{q=1}^{N}d_{lq}x_{til}x_{tjq}\right)^2}\,\right]\\ &+\left[\sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{l=1}^{N}\sum_{q=1}^{N}a_{tilq}x_{(t-1)il}x_{tiq}\right]\Bigg\}. \end{aligned} \tag{1}$$

This formula is subject to:

$$\sum_{i=1}^{N} x_{til} = 1; \qquad \forall t, l. \tag{2}$$

$$\sum_{l=1}^{N} x_{til} = 1; \qquad \forall t, i. \tag{3}$$

$$x_{til} = \begin{cases} 1, & \text{if facility i is assigned to location } l \text{ in period } t \\ 0, & \text{otherwise} \end{cases}. \tag{4}$$

$$M_{ki} - M_{kj} = 1. \tag{5}$$

3.2 | Meta-heuristics

Meta-heuristics are high-level procedures or heuristics aimed at finding, generating or selecting a heuristic. This can provide an appropriate solution to a COP, specifically with incomplete/imperfect information or limited computing capacity. Different meta-heuristics are used to approximate the solution of very large FLPs, for example, SA, Tabu search, CSA, the genetic algorithm, Ant Colony Optimization, particle swarm optimization, and the firefly algorithm (FA). Another class of meta-heuristics is the hybrid meta-heuristic that combines the meta-heuristic with other optimization approaches. The key concept is to combine the high-level algorithms that explore search spaces using different strategies of intensification and diversification. Lee et al[36] and Tayal and Singh[37] applied Hybrid ACO/SA and Hybrid FA/CSA, respectively to solve SDFLP. However recent trends and application in Metaheuristic can studied from Ganesan et al,[38] Gupta and Deep,[39] Vasant et al[40,41] and Zelinka et al.[42]

3.3 | Data envelopment analysis

DEA is a nonstochastic and nonparametric estimation of production frontier based on the actual observations of input-output in the sample. DEA is defined as a mathematical programming formulation applied to observational data in order to estimate the relative efficiency of DMUs. DMUs are compared in terms of multiple inputs and multiple outputs. DEA does not require a specific relationship between inputs and outputs nor fixed weights for the inputs and outputs. The DMUs that lie on the frontier are recognized as efficient, and the remaining as inefficient. The DMU's efficiency is given as the ratio of the weighted sum of its outputs (ie, performance) to the weighted sum of its inputs (ie, resources utilized). DEA requires several inputs and outputs to be considered concurrently to measure DMU efficiency, as given by Equation (6).

$$\text{Efficiency} = \frac{\text{Weighted sum of outputs}}{\text{Weighted sum of inputs}}, \qquad \forall \text{DMUs}, \tag{6}$$

assuming a set of observed DMUs $\{\text{DMU}_\text{j} \mid \text{j} = 1, 2, \dots, n\}$ associated with m inputs $\{x_{ij} \mid \text{i} = 1, 2, \dots, m\}$ and s outputs $\{y_{rj} \mid \text{r} = 1, 2, \dots, s\}$.

DEA finds the most favorable set of weights for each DMU (the set of weights that maximizes the efficiency rating of the DMU without making its own or any other DMU rating greater than one). A basic DEA model can provide important metrics and benchmarks for monitoring and managing actions to improve the comparative performance of entities in a group. A DMU's efficiency score is the distance to this efficient frontier from each DMU. According to this distance, the efficiency scores of inefficient DMUs are calculated and represented as a Pareto ratio. CCR (Charnes, Cooper, and Rhodes) and BCC (Banker, Charnes, and Cooper) are the two common DEA models.

3.4 | Machine learning

The area of ML is typically organized in two main branches: supervised learning and nonsupervised learning. In supervised learning, the ML algorithm receives prelabeled input examples and intends to converge to the best as possible classifier, so one can predict labels for unseen examples with high accuracy. In nonsupervised learning is associated to the process of building up models after analyzing the similarities among input data. For example, the clustering algorithm PCA, k-means attempts to find k representative groups according to the relative distance of points in R^m. The main characteristic of this second type of learning is that algorithms do not have access to labels, therefore the problem is no longer to find a map but instead analyze how points are organized in the input space.[43] Application of intelligent computing can be studied Vasant et al,[44] Panda et al,[45] Abu Zaher et al[46] and integration of SA and clustering algorithm is elaborated in Seifollahi.[47]

4 | META-HEURISTIC-DEA-ML MODEL FOR SOLVING SDFLP

Figure 3 illustrates the steps involved in solving SDFLP using the integrated framework of meta-heuristic, DEA, and ML. In this context, the steps of the proposed solution are as follows:

4.1 | Step 1: SDFLP layout generation

The SDFLP, given by Equation (1) and subject to conditions set in Equations (2) to (5) is simulated to generate the pool of layouts using meta-heuristics—SA (Tayal and Singh[11]), CSA (Tayal and Singh[10]), and Hybrid FA/CSA (Tayal and Singh[37]).

4.2 | Step 2: Identifying criteria for evaluation

The goal of a facility layout design is to streamline the flow of production material, equipment, and manpower in a safe and comfortable environment at minimum costs. The set of criteria that impact facility design are identified from the literature and based on discussions with experts. These are shown in Table 2. All these criteria cannot be mathematically modeled or they may not be congruent to the distance-based formulation given for SDFLP (Equation (1)). Also, the aggregation of these objectives increases the complexity and computation time for solving the SDFLP.

4.3 | Step 3: Dimension reduction

The pool of layouts generated by solving SDFLP (Step 1) needs to be evaluated using the various criteria that affect facility design. The number of factors that affect facility layout are large. Therefore, the unsupervised learning technique is applied for reducing and removing redundant/correlated criteria from the original dataset without disturbing the key objective of the FLP. The steps are as follows.

1. The experts are asked to judge and rank each criterion on a 5-point Likert scale, where 1 is very low, 2 is low, 3 is moderate, 4 is high, and 5 is very high in the scale. The response is obtained from various company experts.
2. Factor analysis is performed to identify a new, smaller set of uncorrelated criteria to replace the original set of correlated criteria. Here, principal components analysis (PCA) is used.

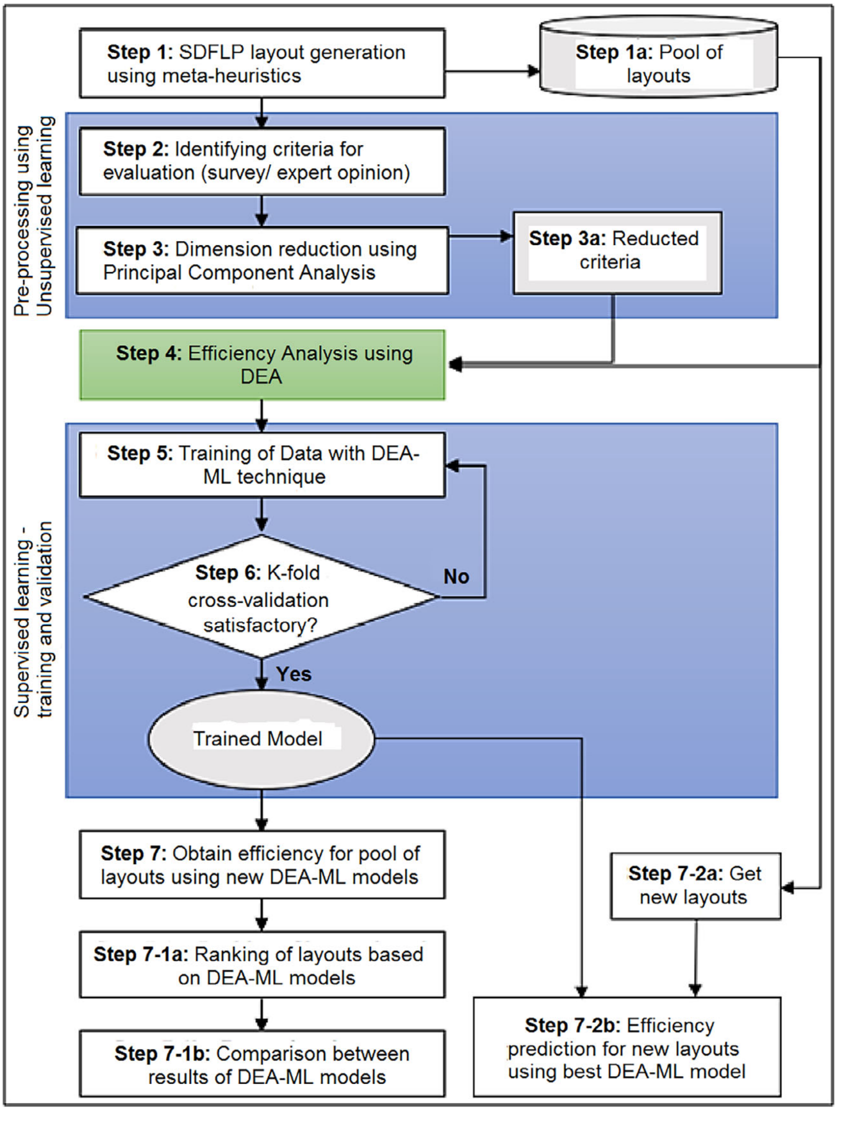

FIGURE 3 Flow chart of the meta-heuristic-data envelopment analysis-machine learning integrated model for solving SDFLP

For the reduced criteria set obtained after factor analysis, the values for each criterion are computed for all of the layouts.

4.4 | Step 4: Efficiency analysis

The set of effective layouts is determined using DEA. For DEA, it is important that all DMUs are functionally similar and homogenous such that the DMUs have the same number and types of inputs and outputs. At the same time, these inputs and outputs need not be congruent. In the case of SDFLP, DMUs are the layouts with inputs and outputs as their criteria.

TABLE 2 List of criteria for stochastic dynamic facility layout problem

Number	Criteria	Description
1	Material handling cost (MHC)	The calculated cost as the product of material flow between two facilities and the distance traveled between them.
2	Flow distance	The sum of volume and distance flow products.
3	Material flow time	The time required to move material between two departments (machines). Unfinished product / waste or finished product can be described as material.
4	Accessibility	The space needed for the material handling path and operator path.
5	Maintenance	To allow uninterrupted operation of manufacturing systems.
6	Waste management	All the activities or actions required from its inception to disposal in order to manage waste.
7	Noise	Acceptable minimum noise in a manufacturing facility for proper operation.
8	Safety	Condition for the workers to avoid harm or any other unwanted outcome.
9	Rearrangement cost (RA_c)	The variable cost of moving a machine over a given period of time from one location to another.
10	Hazardous movement	Material movement with minimal risk and high safety between facilities.
11	Product type	A facility can produce a variety of products.
12	Flexibility	There are two main issues. The first requires the performance of a variety of tasks in the presence of different operating conditions, and the second addresses the flexibility of future expansion.
13	Time period (T)	The time period considered to design the dynamic layout.
14	Number of products (P)	Number of products considered to design the dynamic layout.
15	Facility size (N)	Number of machines to be allocated.

As we know, SDFLP is a COP. Therefore, finding an optimal solution to satisfy all the criteria is not possible. For a given set of inputs, it is possible to generate a number of layouts. As a result, the proposed SDFLP becomes a constant input case for the DEA model. Lovell and Pastor[48] presented three propositions that are adopted for analyzing layout efficiency, given in Figure 4.

The SDFLP has a constant input. Therefore, the BCC model without inputs is adopted for efficiency analysis. A proper number of DMUs is required to identify a true performance frontier.[49] Bowlin[50] suggested that at least two DMUs are required for each input or output measure. The result obtained after solving DEA gives performance frontiers that will be used as the final layouts for design and analysis. At the same time, we are aware that DEA has some disadvantages.

Proposition 1. *An output-oriented CCR model without inputs rates all DMUs as infinitely inefficient.*

Proposition 2. *An output-oriented CCR model with a single constant input coincides with the corresponding BCC model.*

Proposition 3. *An output-oriented BCC model with a single constant input is equivalent to an output- oriented BCC model without inputs*

FIGURE 4 Data envelopment analysis propositions

1. It is possible to rank inefficient DMUs according to their inefficiency values. It is observed that with the increase in input and output the number of efficient DMUs also increases. Ranking DMUs within the efficiency score of 1 is not possible.
2. Curse of dimensionality. For large dataset problems where the number of criteria can be more and/or the number of DMUs is large, an advanced computer with high processing speed and memory is required. Also, it leads to loss of visualization due to the higher number of datasets. To overcome this shortcoming, PCA is applied as discussed in Step 3.
3. To predict the efficiency of the new DMU using the same dataset, DEA-based efficiency analysis cannot be applied without recalculating the efficiency of all DMUs.

4.5 | Step 5: Training of data

For finding the unique efficiency and rankings of layouts, the supervised learning technique is applied. Supervised learning builds an algorithm using input data, a set of features or attributes, which can estimate a specific outcome. It has two stages, learning and inference. In the learning stage, the first step is to describe the data (called a feature vector) and then summarize it into a model. This is the most time-consuming stage, due to the time it takes to converge to a useful model. This model is then used in the inference stage, discussed in Step 7. Figure 5 describes the learning pipeline.

In our research, linear and logistic regression techniques are used. Linear regression is a statistical modeling technique that is used to describe a continuous response variable as a linear function of one or more predictor variables. Logistic regression fits a model that can predict the probability of a binary response that belongs to one class or another.

4.6 | Step 6: *K*-fold cross-validation

After training, a separate dataset is used for testing, which, ideally, should be completely different from the training set. A small predictor error means the algorithm is able to estimate the correct

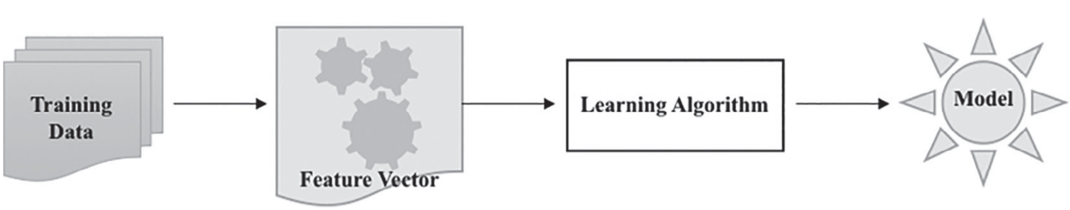

FIGURE 5 Learning pipeline

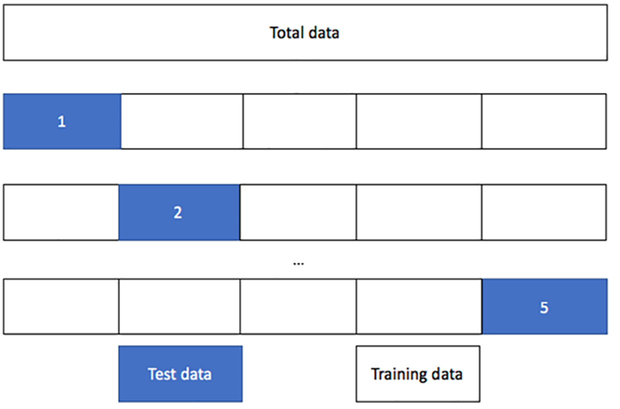

FIGURE 6 K-fold validation, $K = 5$. Data are split randomly into equal size $K = 5$ folds. At each step, one fold is selected as test data and the remaining four are used as training data. The procedure is repeated five times

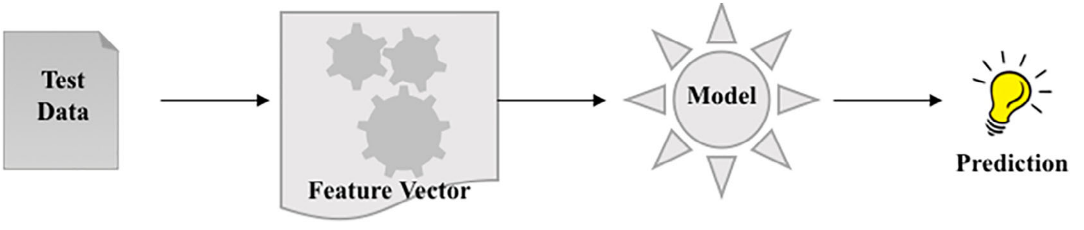

FIGURE 7 Inference pipeline

output (for the test set). To perform this, the commonly used procedure is K-fold cross-validation, which can be studied from Kohavi.[51] Here, the dataset is split into K parts and each split of data is called a fold. The first $K - 1$ fold is used for training (learning algorithm). Then, the data that is held back (validation) is tested. The mean and standard deviations is calculated for each evaluation to obtain the performance score to select the best model (Figure 6).

4.7 | Step 7: Obtain efficiencies for new layouts

From Steps 5 and 6, we get the new DEA-ML models. The efficiencies of all layouts were calculated and are ranked. The mean square errors of each DEA-ML algorithm are compared, and the best algorithm/model is selected.

The efficiency of new layouts is predicted using the best-performing algorithm. This is referred to as inference, the second stage of supervised learning technique, where the model derived in the learning stage is used to produce an intended output on the never-before-seen dataset. Figure 7 describes the inference pipeline.

5 | APPLICATION

The proposed methodology, meta-heuristic-DEA-ML technique, for solving SDFLP is validated using dataset from Moslemipour and Lee.[9] The dataset used is a U-shaped facility with $N = 12$ (Figure 8), $T = 5$, $K = 10$ and Gaussian distribution product demand.

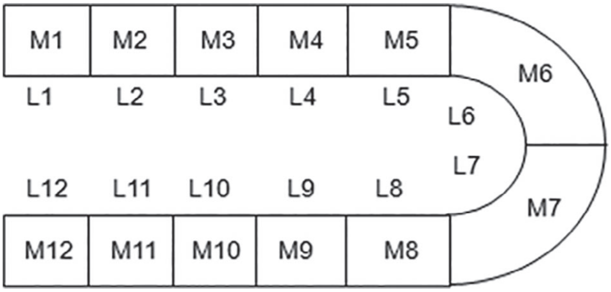

FIGURE 8 U-shaped facility $N = 12$

Equation (1) subject to constraints (2) to (5) is minimized to obtain the pool of layouts, as discussed in Section 3.1, by using meta-heuristics as discussed in Section 3.2. A total of 50 layouts are generated, and an example layout is given below.

$$\text{layout}(s_1) = \begin{Bmatrix} t = 1[6, 2, 7, 4, 9, 11, 3, 10, 5, 1, 12, 8] \\ t = 2[6, 2, 7, 4, 9, 5, 3, 10, 11, 1, 12, 8] \\ t = 3[3, 10, 4, 7, 2, 6, 12, 1, 8, 11, 9, 5] \\ t = 4[11, 10, 3, 9, 1, 8, 12, 6, 7, 5, 4, 2] \\ t = 5[5, 9, 2, 4, 7, 6, 12, 8, 1, 11, 10, 3] \end{Bmatrix}.$$

The layout is represented by a two-dimensional matrix where each row represents a time period, each column represents the facility location, and its element represents the machine number. The MHC and RA_c values for the given layout are 1 243 862 and 35 000, respectively.

There are numerous criteria that influence facility design, given in Table 2. PCA technique is applied to identify reduced criteria/factors. From the survey, the responses for each criterion are obtained and analyzed using SPSS software. Kaiser-Meyer-Olkin test is used to the appropriateness of the factor analysis (value >0.5) and Bartlett's Test of Sphericity (P value $< .05$) is used to examine the hypothesis that criteria are uncorrelated in the survey (Table 3).

Table 4 shows the "Total Variance Explained," which examines the reduction of 12 criteria into smaller dimensional set of five factors. Here, only extracted and rotated values are studied and factors with Eigen values less than 1 can be ignored. As seen in Figure 9, a knee formation occurs between components 5 and 6, which means that five components are sufficient for clustering.

The five factors contribute for 86.271% of the total variance in the data. The component matrix of extracted factor to determine the criteria grouping for the SDFLP under each factor is presented in Table 5. The final set of criteria is shown in Table 6.

MHC and RA_c were already computed in Step 1. For computing the values of the other three criteria, the mathematical formulation is explained below.

TABLE 3 Result of KMO and Bartlett's test

KMO measure of sampling adequacy		**0.744**
Bartlett's test of sphericity	Approx. chi-square	5377.312
	df	66.0
	Sig.	0.000

TABLE 4 Total variance explained

	Initial Eigen values			Extraction sums of squared loadings			Rotation sums of squared loadings		
Component	**Total**	**% of variance**	**Cumulative %**	**Total**	**% of variance**	**Cumulative %**	**Total**	**% of variance**	**Cumulative %**
1	3.488	29.071	29.071	3.488	29.071	29.071	3.011	25.088	25.088
2	2.964	24.702	53.773	2.964	24.702	53.773	2.983	24.855	49.943
3	1.756	14.632	68.405	1.756	14.632	68.405	2.118	17.647	67.59
4	1.107	9.222	77.627	1.107	9.222	77.627	1.201	10.005	77.594
5	1.037	8.645	86.271	1.037	8.645	86.271	1.041	8.677	86.271
6	0.907	7.555	93.827						
7	0.414	3.452	97.278						
8	0.277	2.311	99.589						
9	0.038	0.315	99.904						
10	0.01	0.084	99.988						
11	0.001	0.009	99.998						
12	0	0.002	100						

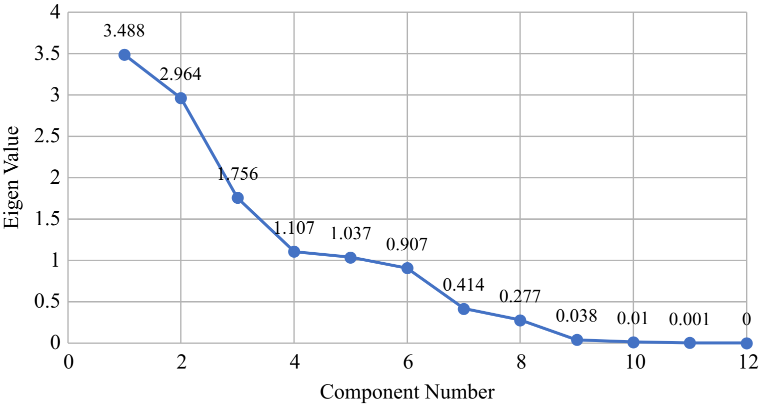

FIGURE 9 Eigen value plot of the 12 components

TABLE 5 Component matrix—extracted factors

Criteria		Component 1	2	3	4	5
Cluster 1	Rearrangement cost	0.222	−0.193	0.117	0.863	0.041
Cluster 2	Waste management	−0.065	−0.106	0.12	−0.24	0.763
	Material flow time	0.034	0.208	−0.089	0.303	0.653
Cluster 3	Maintenance	0.887	−0.264	−0.371	−0.062	0.023
	Flow distance	0.886	−0.265	−0.373	−0.062	0.022
	Accessibility	0.887	−0.263	−0.373	−0.062	0.024
Cluster 4	Noise	0.477	0.849	0.193	−0.02	−0.009
	Safety	0.451	0.852	0.205	−0.005	−0.02
	Hazard management	0.462	0.855	0.195	−0.024	−0.02
Cluster 5	Flexibility	0.418	−0.37	0.567	0.312	−0.113
	Robustness	0.335	−0.382	0.727	−0.17	0.101
	Product demand	0.379	−0.449	0.58	−0.273	−0.021

5.1 | Closeness rating

For SDFLP, the closeness rating function is given by Equation (7), subjected to constraints in Equations (2) to (5),

$$\mathrm{CF}(\pi) = \mathrm{CF}_{iljq} = \sum_{t=1}^{T} \sum_{\mathrm{i}=1}^{\mathrm{N}} \sum_{\mathrm{j}=1}^{\mathrm{N}} r_{ij} \sum_{\mathrm{l}=\mathrm{i}}^{\mathrm{N}} \sum_{\mathrm{q}=1}^{\mathrm{N}} d_{lq} x_{til} x_{tjq}. \tag{7}$$

where r_{ij} is the closeness rating of facilities i and j. CF_{iljq} is the closeness rating function of locating facility i at location l and facility j at location q. Figure 10 shows the closeness rating matrix, where the closeness rating is a function of distance, which means closer the facilities higher will be the rating.

TABLE 6 Final set of criteria

Criteria	Classification	Explanation
Flexibility	**Criteria 1:** Material Handling Cost (MHC)	The facility layout needs to adapt to the fluctuations in product demand, for which the layout should be robust and flexible. These factors affect any manufacturing firm's material handling costs.
Robustness		
Product demand		
Rearrangement Cost	**Criteria 2:** Rearrangement cost (RA_c)	Facilities need to be rearranged as a single layout can be inefficient for all time periods, for this the rearrangement cost is accounted.
Accessibility	**Criteria 3:** Closeness Rating	For ease of handling the material, movement of the operators and facility's maintenance.
Maintenance		
Flow distance		
Hazard	**Criteria 4:** Hazardous movement	For movement of hazardous material and waste, ensuring worker's safety and congeniality of the work place.
Safety		
Noise		
Material Handling Flow Time	**Criteria 5:** Material handling flow time	For material and waste movement.
Waste management		

5.2 | Hazardous movement

For SDFLP, the hazardous movement function is given by Equation (8) subjected to constraints in Equations (2) to (5),

$$\mathrm{HM}(\pi) = \mathrm{HM}_{iljq} = \sum_{t=1}^{T} \sum_{\mathrm{i=1}}^{\mathrm{N}} \sum_{\mathrm{j=1}}^{\mathrm{N}} \mathrm{hm}_{ij} \sum_{\mathrm{l=i}}^{\mathrm{N}} \sum_{\mathrm{q=1}}^{\mathrm{N}} d_{lq} x_{til} x_{tjq} \tag{8}$$

where hm_{ij} is the hazardous movement rating between facilities i and j. It is a time-independent variable. HM_{iljq} is the hazardous movement function between facility i at location l and facility j at location q. The numerical code, “0” means no hazard, “1” is for caution, “2” means prone to minor injury, “3” is warning sign for major injury prone area, and “4” means life-threatening, that is, danger zone, is used. Figure 11 gives the hazardous movement matrix which is a function of distance. For instance, to reduce the risk between facilities whose hazard scores are high the distance between them needs to be maximized.

	1	2	3	4	5	6	7	8	9	10	11	12
1	0	4	8	10	10	6	4	8	10	10	6	4
2	4	0	1	6	2	4	4	1	6	2	4	4
3	8	1	0	4	10	2	8	1	4	10	2	8
4	10	6	4	0	2	4	10	6	4	2	4	10
5	10	2	10	2	0	1	10	2	10	2	1	10
6	6	4	2	4	1	0	6	4	2	4	1	6
7	4	4	8	10	10	6	0	4	4	8	10	10
8	8	1	1	6	2	4	4	0	8	1	1	6
9	10	6	4	4	10	2	4	8	0	10	6	4
10	10	2	10	2	2	4	8	1	10	0	10	2
11	6	4	2	1	1	1	10	1	6	10	0	6
12	4	4	8	10	10	6	10	6	4	2	6	0

FIGURE 10 Adjacency matrix for "N" = 12

5.3 | Material handling flow time

Material and waste movement between two facilities is sequential and calculated by the flow distance divided by the velocity, as given in Equation (9), subjected to constraints in Equations (2) to (5).

$$\mathrm{MHT}(\pi) = \mathrm{MHT}_{iljq} = \sum_{t=1}^{T}\sum_{\mathrm{i}=1}^{\mathrm{N}}\sum_{\mathrm{j}=1}^{\mathrm{N}}\sum_{l=1}^{N}\sum_{q=1}^{N} d_{lq}x_{til}x_{tjq} / \mathrm{velocity} \tag{9}$$

where MHT_{iljq} is the material handling time function between facility i at location l and facility j at location q. In our illustration, the work flow in the process is assumed to be 0.5 units/min. Table 7 gives the values of the criteria for the pool of 50 layouts.

DEA is applied on the 40 layouts from Table 7 to identify the efficient layout, considering the five criteria and their values. The output-oriented BCC model is applied, and an efficiency score is computed, as given in Table 8. It is seen that 21 DMUs have an efficiency of 1.

For unique rankings and prediction, the DEA-ML technique is applied. The supervised ML algorithms were programmed in TensorFlow. TensorFlow was released by Google for ML. It has

	1	2	3	4	5	6	7	8	9	10	11	12
1	0	4	0	0	4	0	0	0	0	4	0	0
2	4	0	1	0	0	0	4	2	4	0	1	0
3	0	1	0	0	0	3	0	3	0	1	0	0
4	0	0	0	0	3	4	2	0	0	0	0	0
5	4	0	0	3	0	0	2	0	4	0	0	3
6	0	0	3	4	0	0	0	4	0	0	3	4
7	0	4	0	2	0	2	0	2	0	4	0	2
8	0	2	3	0	4	0	2	0	0	2	3	0
9	0	4	0	0	4	0	0	0	0	0	4	0
10	4	0	1	0	0	0	4	2	0	0	4	**0**
11	0	1	0	0	0	3	0	3	4	4	0	0
12	0	0	0	0	3	4	2	0	0	0	0	0

FIGURE 11 Hazardous movement for "N" = 12

TABLE 7 Criteria values for the pool of layouts

Layout	Material handling flow time	Rearrangement cost	Closeness rating	Hazardous movement	Material handling cost
#1	4120	46 000	109 500	25 700	1 240 780
#2	4020	35 000	107 910	25 240	1 214 293
#3	4060	40 000	107 710	25 960	1 217 822
#4	3920	46 000	110 530	25 340	1 199 635
#5	4040	36 000	109 930	26 560	1 220 217
#6	4120	35 000	106 060	25 640	1 243 862
#7	4200	27 000	108 070	26 540	1 242 892
#8	4260	29 000	105 960	24 820	1 253 106
#9	4120	34 000	107 540	24 780	1 242 368
#10	4020	47 000	106 370	26 880	1 211 550
#11	4000	47 000	104 810	25 840	1 220 787
#12	4060	48 000	107 160	25 860	1 247 686
#13	4220	37 000	106 770	27 220	1 232 851
#14	4080	31 000	108 800	24 740	1 225 323
#15	4020	41 000	109 600	25 680	1 210 757
#16	4020	44 000	106 080	25 600	1 223 571
#17	4120	43 000	107 810	26 440	1 231 465
#18	4160	47 000	106 930	25 900	1 251 543
#19	4120	46 000	109 500	25 700	1 240 780
#20	4080	22 000	107 790	27 600	1 224 289
#21	4060	35 000	107 380	27 620	1 241 328
#22	4040	46 000	109 310	26 940	1 240 195
#23	4020	47 000	106 370	26 880	1 211 550
#24	4060	39 000	107 130	26 080	1 238 908
#25	4080	45 000	107 870	25 500	1 235 607
#26	4200	45 000	108 640	26 000	1 249 403
#27	4060	43 000	109 740	25 940	1 245 371
#28	4040	43 000	107 600	25 880	1 227 909
#29	3940	29 000	109 390	24 880	1 202 740
#30	4080	44 000	107 200	25 220	1 237 646
#31	4060	40 000	107 710	25 960	1 217 822
#32	4000	44 000	109 930	25 780	1 236 178
#33	4200	37 000	108 710	25 980	1 259 775
#34	4120	36 000	110 170	25 700	1 261 789

(Continues)

TABLE 7 (Continued)

Layout	Material handling flow time	Rearrangement cost	Closeness rating	Hazardous movement	Material handling cost
#35	4020	35 000	107 910	25 240	1 214 293
#36	4100	42 000	109 040	25 780	1 236 755
#37	4080	44 000	109 060	25 820	1 232 708
#38	4100	43 000	109 390	25 020	1 240 858
#39	4140	40 000	105 880	25 840	1 257 739
#40	4120	39 000	109 780	26 280	1 240 628
#41	4200	41 000	108 150	26 000	1 255 274
#42	4100	42 000	110 970	25 660	1 245 548
#43	3920	46 000	110 530	25 340	1 199 938
#44	4120	45 000	107 490	25 480	1 242 866
#45	4040	36 000	109 930	26 560	1 220 217
#46	4080	44 000	109 440	25 460	1 250 412
#47	4220	38 000	106 560	25 800	1 256 637
#48	4080	41 000	109 100	26 360	1 253 514
#49	4120	35 000	106 060	25 640	1 243 862
#50	3920	44 000	110 570	26 740	1 182 795

multiple high-level Application Programming Interfaces (APIs) that are easier to learn and use. They help in managing datasets, estimation, training, evaluation, and inference. A TensorFlow program has two sections:

1. Build the graph for the computation. This is the arrangement of TensorFlow operations in a form of graph nodes.
2. Run the graph.

The actual evaluation of the nodes is done by a TensorFlow session. The computations used for ML are complex and confusing. For easier visualization, debugging, and optimization, there is a tool provided called TensorBoard. In TensorBoard, TensorFlow graphs, plots of metrics on execution of the graph, histograms, etc., can be envisaged. Once TensorBoard is running, the results can be viewed in a web browser with "localhost:6006." Recent developments and applications of TensorFlow in ML can be seen in Abadi et al[52] and at www.tensorflow.org. The TensorFlow source code for DEA-ML includes the steps shown in Figure 12 with the pseudo code given in Figure 13. Here, linear and logistic algorithms are applied for learning to generate a new model using the training/testing data provided in Table 9. The linear regression model is represented in Equation (10), and the cost function for the linear regression model is given by Equation (11). Its tensor graph is shown in Figures 14 and 15, and the plot of cost as a function of training epochs is shown in Figure 16. After training for 1000 epochs, the mean squared loss is 0.000547298. The five

TABLE 8 Efficiency score for the pool of 40 layouts after data envelopment analysis

Layout	Material handling flow time	Rearrangement cost	Closeness rating	Hazardous movement	Material handling cost	Efficiency
#1	4120	46 000	109 500	25 700	1 240 780	1
#2	4020	35 000	107 910	25 240	1 214 293	0.979224727
#3	4060	40 000	107 710	25 960	1 217 822	0.984550705
#4	3920	46 000	110 530	25 340	1 199 635	1
#5	4040	36 000	109 930	26 560	1 220 217	0.999959594
#6	4120	35 000	106 060	25 640	1 243 862	0.987309408
#7	4200	27 000	108 070	26 540	1 242 892	1
#8	4260	29 000	105 960	24 820	1 253 106	1
#9	4120	34 000	107 540	24 780	1 242 368	0.985811394
#10	4020	47 000	106 370	26 880	1 211 550	1
#16	4020	44 000	106 080	25 600	1 223 571	0.979131353
#17	4120	43 000	107 810	26 440	1 231 465	0.996183661
#18	4160	47 000	106 930	25 900	1 251 543	1
#19	4120	46 000	109 500	25 700	1 240 780	1
#20	4080	22 000	107 790	27 600	1 224 289	1
#21	4060	35 000	107 380	27 620	1 241 328	1
#22	4040	46 000	109 310	26 940	1 240 195	1
#23	4020	47 000	106 370	26 880	1 211 550	1
#24	4060	39 000	107 130	26 080	1 238 908	0.988164062
#25	4080	45 000	107 870	25 500	1 235 607	0.990806252
#26	4200	45 000	108 640	26 000	1 249 403	1
#27	4060	43 000	109 740	25 940	1 245 371	1
#28	4040	43 000	107 600	25 880	1 227 909	0.984308329
#29	3940	29 000	109 390	24 880	1 202 740	0.990544739
#30	4080	44 000	107 200	25 220	1 237 646	0.988622363
#31	4060	40 000	107 710	25 960	1 217 822	0.984550705
#32	4000	44 000	109 930	25 780	1 236 178	1
#33	4200	37 000	108 710	25 980	1 259 775	1
#34	4120	36 000	110 170	25 700	1 261 789	1
#35	4020	35 000	107 910	25 240	1 214 293	0.979224727
#36	4100	42 000	109 040	25 780	1 236 755	0.994459112
#37	4080	44 000	109 060	25 820	1 232 708	0.994831287
#38	4100	43 000	109 390	25 020	1 240 858	0.996905829
#39	4140	40 000	105 880	25 840	1 257 739	0.999967435

(Continues)

TABLE 8 (Continued)

Layout	Material handling flow time	Rearrangement cost	Closeness rating	Hazardous movement	Material handling cost	Efficiency
#40	4120	39 000	109 780	26 280	1 240 628	1
#46	4080	44 000	109 440	25 460	1 250 412	1
#47	4220	38 000	106 560	25 800	1 256 637	1
#48	4080	41 000	109 100	26 360	1 253 514	1
#49	4120	35 000	106 060	25 640	1 243 862	0.987309408
#50	3920	44 000	110 570	26 740	1 182 795	1

1. Loading the training and testing data into TensorFlow
   - 40 layouts train test data, Input: Material Handling Cost, Rearrangement Cost, Material Handling Flow Time, Closeness Rating, Hazardous Movement, Output: Efficiency (from DEA)
2. Constructing the learning model classifier
   - Learning algorithm: Linear Regression, Logistic Regression
3. Start the TensorFlow session for training the model and cross-validating using K-fold method split = (8)
4. Accuracy of the model
5. Predicting for new sample data

FIGURE 12 Data envelopment analysis-machine learning tensor flow steps

weights, because of the five criteria, are weight_1 = 0.20315343, weight_2 = 0.04342238, weight_3 = 0.02651982, weight_4 = −0.14450487, and weight_5 = 0.04755654. The bias is b = 0.92481017.

$$\text{Y\_predicted} = [\text{weight}][X] + b \tag{10}$$

$$\text{cost} = (Y - \text{Y\_predicted})^2, \tag{11}$$

where X is a vector holding the value of each criteria for each layout and Y is the layout efficiency from the DEA.

The logistic regression model is represented in Equation (12), and the cost function for the linear regression model is given by Equation (13). Its tensor graph is shown in Figure 15, and the plot of cost as a function of training epochs is shown in Figure 16. After training for 1000 epochs, the mean squared loss is 5.52048e-05. The five weights are weight_1 = 0.64860719, weight_2 = 0.47263116, weight_3 = 0.4062393, weight_4 = 1.58002031, and weight_5 = 1.57390189. The bias is b = 2.07157826.

$$\text{Y\_predicted} = \text{Sigmoid}([\text{weight}][X] + b) \tag{12}$$

$$\text{cost} = \text{mean}\left(\sum((-Y * \log(\text{Y\_predicted})) + (1 - Y) * \log(1 - \text{Y\_predicted}))\right) \tag{13}$$

```
1.
<Import the tensor libraries>
Read the training and test data from excel, csv, text files
      Storing the data as array, data = np.asarray(…)
Classify the data as input and output and store in variables
Define place holder for X (criteria) and Y (efficiency)
      X = tf.placeholder(tf.float32, shape=[None, 5], name=”x”)
      Y = tf.placeholder(tf.float32, shape=[None, 1], name=”Y”)
Define the weights and bias variables as appropriate
      w = tf.Variable(tf.random_normal(shape=[]nDim,1), name=”weight”)
      b = tf.Variables(tf.random_normal([1]), name=”bias”)
2.
Initialize the variables
      Init = tf.global_variables_initializer()
Define Y predict function: linear or logistic regression
Define the cost function for linear and logistic regression
Define the learning rate and epoch
      learningRate = 0.01
      trainingEpoch = 1000
Use gradient descent to minimize the cost
      optimize =
      tf.train.GradientDescentOptimizer(learningRate).minimize(cost)
3.
Start the tensor session and initialize
      sess = tf.Session()
      sess.run(init)
for epoch in range(trainingEpoch):
      for k.split(data…8) for k-fold cross validation:
            Run the optimizer
            Calculate the cost function
4.
            Run the predictor on the test data
            Calculate the mean square error

5.
Read the layout data for prediction
Use the new model to predict the efficiency. Of the 50-layout pool
Close the tensor session
```

FIGURE 13 Data envelopment analysis-machine learning tensor flow pseudo code

For the 40 layouts, the efficiency rankings using linear regression and logistic regression models are captured in Table 9.

In Table 9, Layout#26, given below, is ranked 1 as per the models.

$$\text{layout}(s_{26}) = \left\{ \begin{array}{l} t = 1[6, 7, 2, 4, 9,11,5, 3,10,1, 12, 8] \\ t = 2[8,12,11,10,3, 9, 5, 4, 2, 7, 6, 1] \\ t = 3[12, 8, 4, 2, 6, 7, 11,10,3, 9, 5, 1] \\ t = 4[8, 1, 12,11,9, 5, 3, 10, 2, 4, 6, 7] \\ t = 5[3,10,11,2, 4, 7, 6, 1, 8,12,9, 5] \end{array} \right\}.$$

Table 10 shows that the mean squared error (MSE) for logistic regression is less than that of linear regression. Also, the efficiency scores and ranking obtained from DEA map significantly with those obtained from DEA-ML (logistic regression). The efficiency of the remaining 10 layouts from the pool of layouts is predicted using the logistic regression model, given in Table 11.

TABLE 9 Comparison of rankings among DEA-BCC, DEA-ML (linear regression model), and DEA-ML (logistic regression model)

Layout	Efficiency DEA-BCC	Rank DEA-BCC	Efficiency DEA-ML (linear regression)	Rank DEA-ML (linear regression)	Efficiency DEA-ML (logistic regression)	Rank DEA-ML (logistic regression)
#1	1	1	1.02403784	5	0.99616987	3
#2	0.979224727	38	0.97717577	27	0.98179501	37
#3	0.984550705	35	1.00898623	14	0.9892652	33
#4	1	1	0.9837485	23	0.98768741	35
#5	0.999959594	23	1.00651228	18	0.98950231	32
#6	0.987309408	32	0.97347832	31	0.99084598	26
#7	1	1	1.03087699	2	0.99018985	30
#8	1	1	1.02098942	7	0.99052703	28
#9	0.985811394	34	0.97689581	29	0.98991126	31
#10	1	1	1.00838876	16	0.99161106	24
#16	0.979131353	40	0.96708047	37	0.99032462	29
#17	0.996183661	25	1.03049171	3	0.99433511	17
#18	1	1	1.01380384	11	0.99680871	2
#19	1	1	1.02403784	5	0.99616987	3
#20	1	1	0.99184799	21	0.97893435	39
#21	1	1	0.96980035	35	0.992607	21
#22	1	1	0.98738593	22	0.99615461	5
#23	1	1	1.00838876	16	0.99161106	24
#24	0.988164062	31	0.96560073	38	0.99228716	22
#25	0.990806252	28	0.99443966	20	0.99433303	18
#26	1	1	1.05093563	1	0.99698728	1
#27	1	1	0.97820312	26	0.99554402	10
#28	0.984308329	37	0.98255891	24	0.99222434	23
#29	0.990544739	29	0.94665223	40	0.96455085	40
#30	0.988622363	30	0.98157936	25	0.99367416	20
#31	0.984550705	35	1.00898623	14	0.9892652	33
#32	1	1	0.96086454	39	0.99430943	19
#33	1	1	1.01763844	10	0.99604034	6
#34	1	1	0.97203767	33	0.99562424	9
#35	0.979224727	38	0.97717577	27	0.98179501	37
#36	0.994459112	27	1.00962818	12	0.99438226	16
#37	0.994831287	26	1.00924873	13	0.9944331	15

(Continues)

TABLE 9 (Continued)

Layout	Efficiency DEA-BCC	Rank DEA-BCC	Efficiency DEA-ML (linear regression)	Rank DEA-ML (linear regression)	Efficiency DEA-ML (logistic regression)	Rank DEA-ML (logistic regression)
#38	0.996905829	24	0.99942046	19	0.99471951	13
#39	0.999967435	22	0.96976149	36	0.99511731	12
#40	1	1	1.0188266	9	0.99471468	14
#46	1	1	0.97622621	30	0.99600154	7
#47	1	1	1.02034509	8	0.99538213	11
#48	1	1	0.97095513	34	0.99592638	8
#49	0.987309408	32	0.97347832	31	0.99084598	26
#50	1	1	1.02406955	4	0.98413104	36

Abbreviations: DEA-BCC, data envelopment analysis-Banker, Charnes, and Cooper; DEA-ML, data envelopment analysis-machine learning.

5.4 | Limitations of the study

Although the obtained findings provide positive trends in terms of the proposed methodology, it is also possible to explain some about limitations, too. In this sense, the meta-heuristic-DEA-ML solution for measuring and predicting layout efficiency for SDFLP is a system of different components, which should be carefully used according to synergy among each of different components. So, it may be give better opportunity in terms of being practical if further research

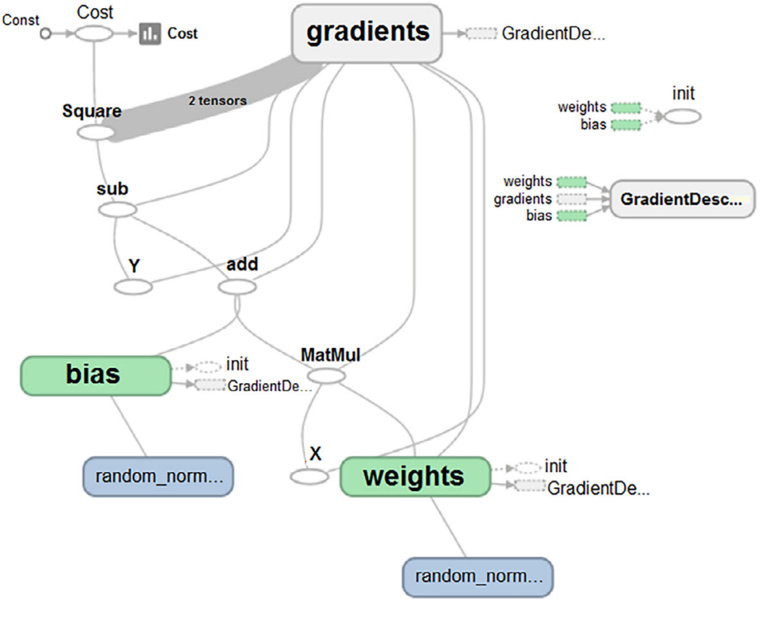

FIGURE 14 Tensor graph for linear regression

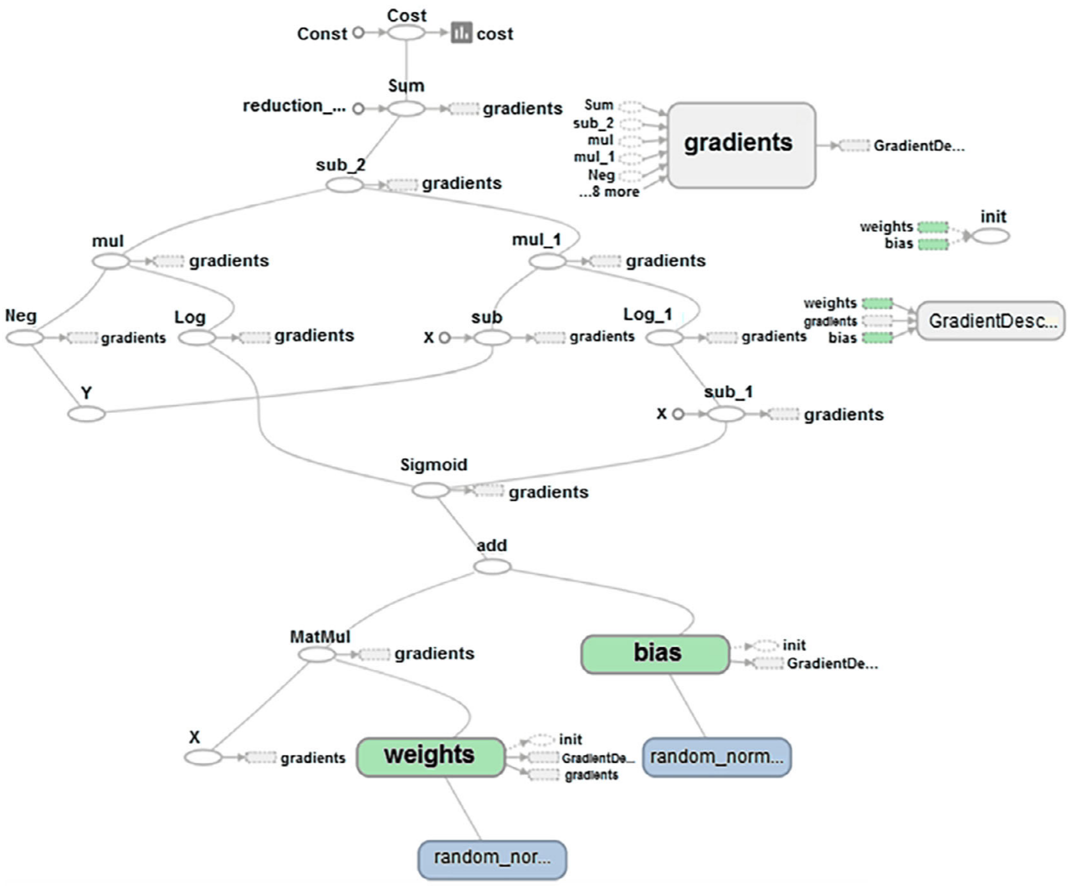

FIGURE 15 Tensor graph for logistic regression

works provide a solution with simpler algorithmic structure. On the other hand, there is the deep learning, which is a trendy and more advanced form of ML. The solution proposed in this study may be overcome by deep learning-based techniques and the approach of deep learning may be more effective in terms of analyzing more and more data at the same time. However, of course applying deep learning will require use of stronger computer systems so applicability in this manner may be an advantage for the meta-heuristic-DEA-ML solution of

TABLE 10 Mean squared error (MSE) for the learning algorithms

Training epoch	MSE DEA-ML (linear regression)	MSE DEA-ML (logistic regression)
10	0.0800474	0.0214715
100	0.0123861	0.000708778
500	0.00710666	8.76759e-05
1000	0.000547298	5.52048e-05

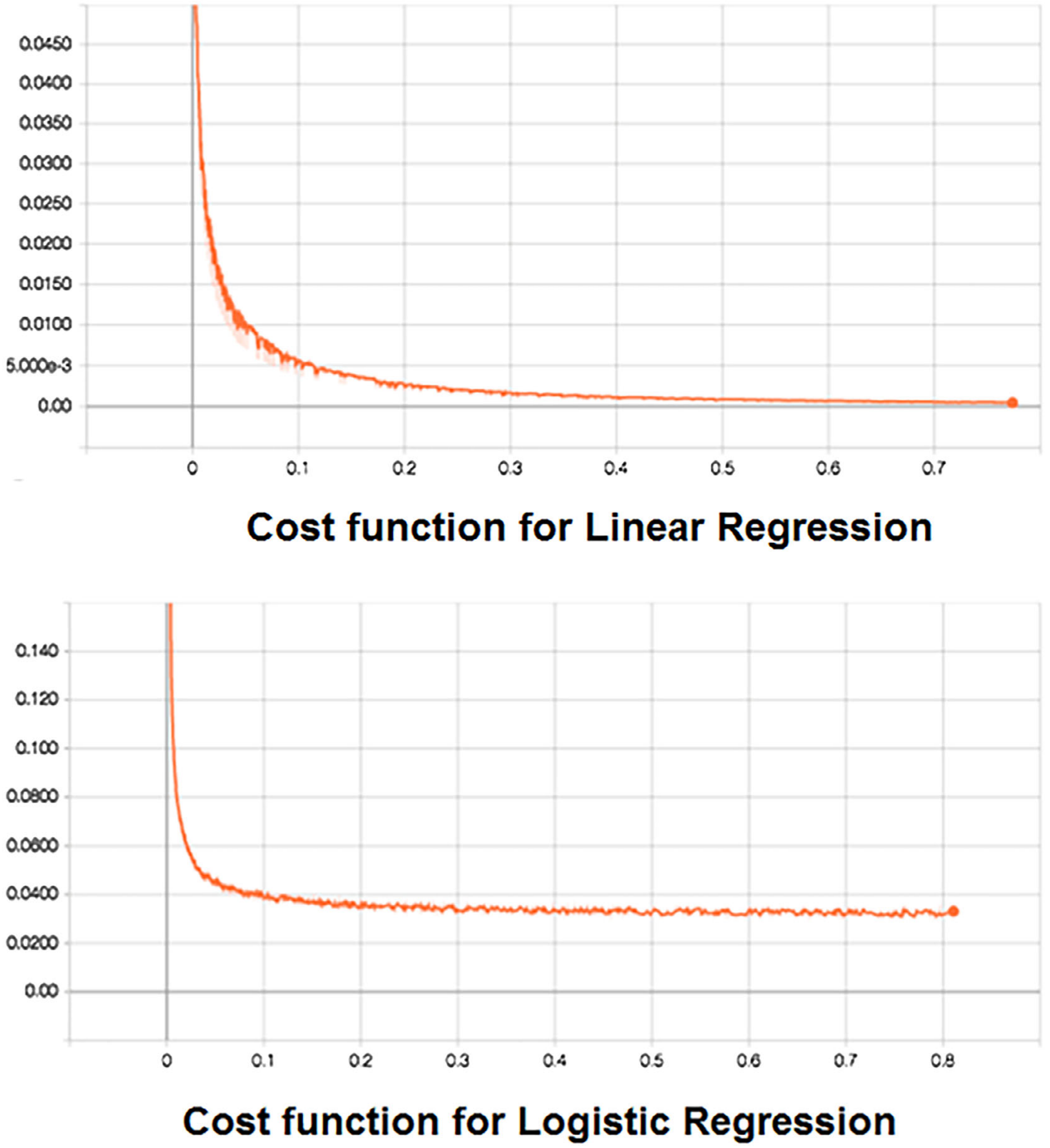

FIGURE 16 Cost function of learning algorithms

this study. Additionally, SDFLP is a problem, which can be improved and changed according to different conditions so it is a paradox to express that the solution of this study will always solve every kind of SDFLP problem. That limitation is important since it gives an opportunity to develop alternative systems and analyze alternative SDFLP problems always. As a final limitation, it can be expressed that the intelligent optimization and hybrid systems development for ML and intelligent systems are all active fields so that new algorithms or solution ways may affect actuality of the solution of this study. That is currently a limitation but of course also an opportunity to keep the associated fields alive and continue to further works.

TABLE 11 Prediction efficiency of new layouts

Layout	Efficiency DEA-ML (logistic regression)
#41	0.99645823
#42	0.99586135
#43	0.98776102
#44	0.99530679
#45	0.98950225
#11	0.99055356
#12	0.9961229
#13	0.99365234
#14	0.98334748
#15	0.98870343

Abbreviations: DEA-ML, data envelopment analysis-machine learning.

6 | CONCLUSION AND FUTURE SCOPE

The research explores the potential of an integrated approach of meta-heuristic-DEA-ML for measuring and predicting layout efficiency for SDFLP with conflicting criteria. For SDFLP, traditional efficiency assessment approaches are not effective because they do not account for the complete range of criteria and data such as social, economic, political, environmental, personal, international standards, and risk. The aim of this paper is to demonstrate the practical and analytical applications of the integrated meta-heuristic-DEA-ML methodology to optimize the FLP. Also, the subjectivity of decisions, due to an expert's opinion in analyzing various criteria related to facility design, is handled by the integration of ML with DEA. Furthermore, in the presence of uncertainty, the layout considered here takes into account multiple products in multiple time periods.

Data and criteria are collected from experts across manufacturing industries in the proposed methodology. When resolving SDFLP, all of the factors are not considered relevant, as some can be redundant, conflicting or less significant. To help derive a manageable set of factors, the heterogeneity properties of PCA are used. Meta-heuristic techniques are used to generate pools of layouts. Then, DEA is applied on the conflicting criteria to compute the efficiency scores of these layouts. Finally, these data are used for supervised learning to get a trained and validated model, which is then used for the efficiency assessments and ranking of layouts. Two supervised ML models are considered for evaluation (linear and logistic regression). The logistic regression learning model shows the smallest MSE. Therefore, it is used for predicting the efficiency of new layouts.

The future research direction is to collect large datasets from manufacturing industry based on different criteria such as finance, policy, pollution, energy, and sustainability. The proposed framework can be analyzed integrating other ML techniques, such as deep learning, with DEA for predicting the efficiency of layouts. Finally, this framework can be used to investigate, evaluate and predict the impact on efficiency of the layout because of the change in criteria or strategies. This will facilitate the managers to have a portfolio choice strategy based on the varying requirements of the firm.

ACKNOWLEDGMENTS

The authors thank the organizers of ICO'2018, October 4 to 5, 2018, Hard Rock Hotel, Pattaya, Thailand, for the opportunity to presents their research results.

ORCID

Utku Kose https://orcid.org/0000-0002-9652-6415
Anand Nayyar https://orcid.org/0000-0002-9821-6146
José Antonio Marmolejo Saucedo https://orcid.org/0000-0002-8539-9828

REFERENCES

1. Tompkins JA, White JA, Tanchoco JMA. *Facilities Planning*. New York, NY: Willey; 2003.
2. Rosenblatt MJ, Lee H. A robustness approach to facilities design. *Int J Prod Res*. 1987;25(4):479-486.
3. Kusiak A, Heragu SS. The facility layout problem. *Eur J Oper Res*. 1987;29(3):229-251.
4. Meller RD, Gau KY. The facility layout problem: recent and emerging trends and perspectives. *J Manuf Syst*. 1996;15:351-366.
5. Koopmans TCS, Beckman M. Assignment problem and the location of economic activities. *Econometric*. 1957;25(1):53-76.
6. Shore RH, Tompkins JA. Flexible facilities design. *AIIE Trans*. 1980;12(2):200-205.
7. Rosenblatt MJ, Kropp DH. The single period stochastic plan layout problem. *IIE Trans*. 1992;24(2):169-176.
8. Palekar US, Batta R, Bosch RM, Elhence S. Modeling uncertainties in plant layout problems. *Eur J Oper Res*. 1992;63:347-359.
9. Moslemipour G, Lee TS. Intelligent design of a dynamic machine layout in uncertain environment of flexible manufacturing systems. *J Intell Manuf*. 2011;23(5):1849-1860.
10. Tayal A, Singh SP. Chaotic simulated annealing for solving stochastic dynamic facility layout problem. *J Int Manag Stud*. 2014;14(2):67-74.
11. Tayal A, Singh SP. Analysis of simulated annealing cooling schemas for design of optimal flexible layout under uncertain dynamic product demand. *Int J Oper Res*. 2019;34:85.
12. Tayal A, Singh SP. Analysing the effect of chaos functions in solving stochastic dynamic facility layout problem using CSA. *Advances in Intelligent Systems and Computing*. Singapore: Springer; 2016:99-108.
13. Charnes A, Cooper WW, Rhodes E. Measuring the efficiency of decision making units. *Eur J Oper Res*. 1978;2(6):429-444.
14. Banker RD, Charnes A, Cooper WW. Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Manag Sci*. 1984;30(9):1078-1092.
15. Charnes A, Cooper WW, Golany B, Seiford L, Stutz J. Foundations of data envelopment analysis for Pareto-Koopmans efficient empirical production functions. *J Econ*. 1985;30(1–2):91-107.
16. Tone K. A slacks-based measure of efficiency in data envelopment analysis. *Eur J Oper Res*. 2001;130(3):498-509.
17. Emrouznejad A, Parker BR, Tavares G. Evaluation of research in efficiency and productivity: a survey and analysis of the first 30 years of scholarly literature in DEA. *Socioecon Plann Sci*. 2008;42(3):151-157.
18. Yang T, Kuo C. A hierarchical AHP/DEA methodology for the facilities layout design problem. *Eur J Oper Res*. 2003;147:128-136.
19. Azadeh A, Moghaddam M, Asadzadeh M, Negahban A. Optimization of facility layout design with ambiguity by an efficient fuzzy multivariate approach. *Int J Adv Manuf Technol*. 2015;84(1–4):565-579.
20. Tayal A, Singh SP. Integrated SA-DEA-TOPSIS based solution approach for multi objective stochastic dynamic facility layout problem. *Int J Bus Syst Res*. 2017a;11:82.
21. Tayal A., Gunasekaran A., Singh S.P., Dubey R. Papadopoulos T. Formulating and solving sustainable stochastic dynamic facility layout problem: A key to sustainable operations.. *Ann Oper Res*. 2017;253(1):621-655.
22. Athanassopoulos AD, Curram SP. A comparison of data envelopment analysis and artificial neural networks as tools for assessing. *J Oper Res Soc*. 1996;47(8):1000-1016.
23. Liu HH, Chen TY, Chiu YH, Kuo FH. A comparison of three-stage DEA and artificial neural network on the operational efficiency of semi-conductor firms in Taiwan. *Mod Econ*. 2013;4(1):20-31.

24. Liao H, Wang B, Weyman-Jones T. Neural network based models for efficiency frontier analysis: an application to East Asian economies' growth decomposition. *Glob. Econ. Rev.* 2007;36(4):361-384.
25. Santin D. On the approximation of production functions: a comparison of artificial neural networks frontiers and efficiency techniques. *Appl Econ Lett*. 2008;15(8):597-600.
26. Santín D, Delgado FJ, Valiño A. The measurement of technical efficiency: a neural network approach. *Appl Econ*. 2004;36(6):627-635.
27. Wang S. Adaptive non-parametric efficiency frontier analysis: a neural-network based model. *Comput Oper Res*. 2003;30(2):279-296.
28. Azadeh A, Amalnick MS, Ghaderi SF, Asadzadeh SM. An integrated DEA PCA numerical taxonomy approach for energy efficiency assessment and consumption optimization in energy intensive manufacturing sectors. *Energy Policy*. 2007;35:3792-3806.
29. Mostafa M. Modeling the efficiency of top Arab banks: a DEA-neural network approach. *Expert Syst Appl*. 2009;36(1):309-320.
30. Wu D. Supplier selection: a hybrid model using DEA, decision tree and neural network. *Expert Syst Appl*. 2009;36(5):9105-9112.
31. Kwon H. Performance modeling of mobile phone providers: a DEA-ANN combined approach. *Benchmarking*. 2014;21(6):1120-1144.
32. Hanafizadeh P, Khedmatgozar HR, Emrouznejad A, Derakhshan M. Neural network DEA for measuring the efficiency of mutual funds. *Int J Appl Decis Sci*. 2014;7(3):255-269.
33. Misiunas N, Oztekin A, Chen Y, Chandra K. DEANN: A healthcare analytic methodology of data envelopment analysis and artificial neural networks for the prediction of organ recipient functional status. *Omega*. 2016;58:46-54.
34. Kim B, Lee H, Kang P. Integrating cluster validity indices based on data envelopment analysis. *Appl Soft Comput*. 2018;64:94-108.
35. De Clercq, Djavan, Devansh Jalota, Ruoxi Shang, Kunyi Ni, Zhuxin Zhang, Areeb Khan, Zongguo Wen, Luis Caicedo, and Kai Yuan. (2019) "Machine learning powered software for accurate prediction of biogas production: a case study on industrial-scale Chinese production data." *J Clean Prod*. 218:390–399.
36. Lee, T. S., Moslemipour, G., Ting, T. O., Rilling, D., (2012). A novel hybrid ACO/SA approach to solve stochastic dynamic facility layout problem (SDFLP), ICIC (3), *Commun Comput Inf Sci*, Vol. 304, 100–108, Springer.
37. Tayal, A. and Singh, S. P. (2017bb). Designing flexible stochastic dynamic layout: an integrated firefly and chaotic simulated annealing based approach. *Glob J Flex Syst Manag*, 18 (2), 89–98.
38. Ganesan T, Vasant P, Elamvazuthi I. *Advances in Metaheuristics: Applications in Engineering Systems*. Boca Raton, FL: CRC Press; 2016.
39. Gupta S, Deep K. Random walk grey wolf optimizer for constrained engineering optimization problems. *Comput Intell*. 2018;34(4):1025-1045.
40. Vasant P, Kose U, Watada J. Metaheuristic techniques in enhancing the efficiency and performance of thermo-electric cooling devices. *Energies*. 2017;10(11):1703.
41. Vasant P, Marmolejo JA, Litvinchev I, Aguilar RR. Nature-inspired meta-heuristics approaches for charging plug-in hybrid electric vehicle. *Wirel Netw*. 2019;1-14. (In Press) https://doi.org/10.1007/s11276-019-01993-w.
42. Zelinka I, Tomaszek L, Vasant P, Dao TT, Hoang DV. A novel approach on evolutionary dynamics analysis–a progress report. *J Comput Sci*. 2018;25:437-445.
43. Nguyen HL, Woon YK, Ng WK. A survey on data stream clustering and classification. *Knowl Inf Syst*. 2015;45(3):535-569.
44. Vasant P, Zelinka I, Weber G-W. *Intelligent Computing & Optimization*. Conference Proceedings of the ICO 2018. Cham, Switzerland: Springer; 2018.
45. Panda MR, Das PK, Dutta S, Pradhan SK. Optimal path planning for mobile robots using oppositional invasive weed optimization. *Comput Intell*. 2018;34(4):1072-1100.
46. Abu Zaher A, Berretta R, Noman N, Moscato P. An adaptive memetic algorithm for feature selection using proximity graphs. *Comput Intell*. 2019;35(1):156-183.
47. Seifollahi S, Bagirov A, Zare Borzeshi E, Piccardi M. A simulated annealing-based maximum-margin clustering algorithm. *Comput Intell*. 2019;35(1):23-41.
48. Lovell CK, Pastor JT. Radial DEA models without inputs or without outputs. *Eur J Oper Res*. 1999;118(1):46-51.

49. Zhang Y, Bartels R. The effect of sample size on the mean efficiency in DEA with an application to electricity distribution in Australia, Sweden and New Zealand. *J Prod Anal*. 1998;9(3):187-204.
50. Bowlin WF. Evaluating the efficiency of US Air Force real-property maintenance activities. *J Oper Res Soc*. 1987;38(2):127-135.
51. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. *Int Joint Conf Artif Intell*. 1995;14(2):1137-1145.
52. Abadi M, Barham P, Chen J, et al. TensorFlow: a system for large-scale machine learning. Paper presented at: 12th USENIX Symposium on Operating Systems Design and Implementation; 2016; vol. 16:265–283.

How to cite this article: Tayal A, Kose U, Solanki A, Nayyar A, Saucedo JAM. Efficiency analysis for stochastic dynamic facility layout problem using meta-heuristic, data envelopment analysis and machine learning. *Computational Intelligence*. 2019;1–31. https://doi.org/10.1111/coin.12251