Asian Journal of Innovation and Policy (2024) 13.1:001-028
DOI: https://doi.org/10.7545/ajip.2024.13.1.001

Integrating Machine Learning with Data Envelopment Analysis for Enhanced R&D Efficiency & Optimizing Resource Allocation in the Specialized Field

Seokki Cha*, Kyunghwan Park**

Abstract Enhancing the efficiency of research and development (R&D) is crucial for organizations to remain competitive and generate innovative solutions. Data Envelopment Analysis (DEA) has emerged as a powerful tool for evaluating R&D efficiency. However, traditional DEA models heavily rely on the selection of input and output variables, which can limit their effectiveness. To overcome this dependency and improve the robustness of DEA, this study proposes a novel methodology that integrates machine learning techniques with DEA for determining the most suitable input and output variables. The proposed approach is particularly relevant for specialized R&D fields, such as Radiation Emergency Medicine (REM). REM is a critical domain that deals with the medical and public health consequences of nuclear emergencies. The selection of REM as the focus of this study is motivated by several factors, including the unique challenges posed by the field, the potential for significant societal impact, and the need for efficient resource allocation in emergency situations. By leveraging machine learning algorithms, such as Support Vector Machines (SVM), the proposed methodology aims to identify the most relevant input and output variables for DEA in the context of REM. The integration of machine learning enables the DEA model to capture complex relationships and non-linearities in the data, leading to more accurate and reliable efficiency assessments. The effectiveness of the proposed methodology is demonstrated through a comprehensive evaluation using real-world REM data. The results highlight the superior performance of the machine learning-integrated DEA approach compared to traditional DEA models. This study contributes to the advancement of R&D efficiency assessment in specialized fields and provides valuable insights for decision-makers in REM and other critical domains.

Submitted, February 20, 2024; 1st Revised, April 23, 2024; Accepted, May 8, 2024
* Corresponding, Senior Researcher, R&D Strategy Center, Korea Institute of Science and Technology Information (KISTI), Daejeon, Republic of Korea; sc04@kisti.re.kr
** 2nd Author, Team Learder, REM Policy Team, Korea Institute of Radiological and Medical Sciences (KIRAMS), Seoul, Republic of Korea; pkhwan@kirams.re.kr


This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Keywords Data Envelopment Analysis; Machine Learning; Radiation Emergency Medicine; R&D Efficiency; R&D Management; Specialized Research Domain; Resource Allocation; Input and Output Parameters

I. Introduction

The dynamic advancement of science and technology has emerged as a central determinant of competitiveness among nations worldwide(Malik, 2012). In line with this trend, various countries are significantly increasing their annual budgets in the field of science and technology(Inekwe, 2015). However, mere increases in research and development (R&D) budgets are insufficient to drive progress(Wang et al., 2013). This stark reality necessitates diverse methodologies for efficient utilization of R&D resources, particularly in fields with strong public interest.

In addition to prevailing methodologies, this study incorporates two primary techniques: data envelopment analysis (DEA) and machine learning. The DEA model, traditionally employed for efficiency assessment, has been augmented by the innovative application of machine learning algorithms within this investigation. Introduced as a pioneering technique, machine learning serves as a novel method, meticulously crafted to refine conventional DEA models through the optimization of multifaceted factors, thereby culminating in more accurate outcomes. The amalgamation of machine learning and DEA engenders an unprecedented synergy that furnishes dynamic and exhaustive solutions for the evaluation of R&D resources, with particular emphasis on the field of Radiation Emergency Medicine (REM). This fusion of methodologies not only enhances the evaluative process but also emphasizes the pivotal role of machine learning in realizing optimal efficiency within the framework of this research.

This study aims to employ the DEA model as a method to achieve maximum efficiency of science and technology R&D resources. DEA is a model that provides dynamic, aggregate, and comparative results for evaluating manufacturers' supply chains and organizational performance, organizational productivity, etc. (Abdullah et al., 2023) and can be implemented computationally through Python programming. Recognizing the limitation of conventional DEA models, where efficiency evaluation can vary significantly depending on input and output factors (Chen et al., 2021), we sought to design a procedure using machine learning algorithms.

In previous studies, there has also been a study that optimizes the DMU of DEA with a clustering algorithm by incorporating machine learning, which is frequently used for optimization(Mirmozaffari et al., 2023). The integration of Data Envelopment Analysis (DEA) with machine learning represents a

significant stride in R&D efficiency evaluation. This synergy allows for a nuanced understanding of the qualitative and quantitative impact of employing these methodologies, particularly in the field of Radiation Emergency Medicine (REM). By focusing on optimizing R&D efficiency, our novel approach not only refines conventional methodologies but also showcases the broader potential impact, bridging technological innovation with critical public domains. This connection elucidates the significant societal implications and underlines the transformative possibilities unlocked by this research, providing a comprehensive perspective on the alignment of technology with pressing societal needs.

The choice of the DEA model for evaluating R&D efficiency in this study was based on careful consideration of the specific characteristics of the data used. Unlike typical R&D projects where there might be a significant time lag between inputs and outputs, the data employed in this research consists of annual input and output measurements for each research task. This unique feature allows for the application of the DEA model without the need to account for temporal disparities. By utilizing data that captures inputs and outputs within the same annual timeframe, we ensure the suitability and accuracy of the DEA approach for assessing R&D efficiency in this particular context.

Furthermore, the use of annual input-output data in this study facilitates the development of a more universally applicable methodology. By demonstrating the effectiveness of the DEA model in evaluating R&D efficiency within a specific temporal framework, we aim to provide a standardized approach that can be readily adopted by researchers and practitioners across various fields. This emphasis on generalizability enhances the intellectual understanding of readers and promotes the broader applicability of our findings, ultimately contributing to the advancement of R&D efficiency assessment practices.

This study aims to derive the input and output parameters of DEA considering the specificity of R&D through this innovative machine learning approach. To this end, this study derives the most suitable factor for DEA by optimizing R&D efficiency in the field of REM (Cha et al., 2020) because, after the Fukushima nuclear power plant accident in 2011, the field of REM rose to prominence (Cha et al., 2020; Tatsuzaki et al., 2022; Coleman et al., 2012). The unique characteristics of a non-profit field that fuses nuclear power and medical fields with a focus on public interest make it suitable for this study.

Thinking that machine learning-based DEA results could be more objectively compared with quantitative benchmarks rather than solely relying on human qualitative assessments, a new model was devised that could more precisely measure R&D efficiency. Preliminary interviews with experienced researchers

and R&D managers in the field of REM were conducted to select the most relevant input and output factors. Subsequently, this study employed existing DEA, machine learning-integrated DEA, and quantitative benchmarks derived from historical data trends and peer-reviewed R&D efficiency metrics for a comparative evaluation. Through Pearson correlation analysis, it was confirmed that the machine learning-integrated DEA displayed a higher correlation with these quantitative benchmarks than with conventional DEA, illustrating a more objective and reliable measure of efficiency.

The main findings affirm the enhanced capability of DEA when being integrated with machine learning, suggesting novel pathways to refine the accuracy of efficiency assessments across distinct research fields. This study particularly validates the application of machine learning as an innovative technique to transcend the limitations traditionally associated with DEA. By merging machine learning with the established DEA methodology, this study substantiated a targeted approach for evaluating R&D efficiency within specialized areas of research and development. Hence, the study can contribute significantly to the discourse on R&D efficiency, offering evidence-based insights that are of strategic value to policymakers and practitioners alike.

II. Theoretical Backgrounds

1. Research Question

In this study, we seek to address the research question of 'how integration of machine learning algorithms with Data Envelopment Analysis (DEA) models can improve the evaluation of relative efficiency among decision units (DMUs),' taking into account the particularities of specific R&D fields. We explore how the existing DEA model can be improved by selecting input and output parameters to accurately reflect the unique characteristics and constraints of this field. This question guides us to explore innovative methodologies for optimizing R&D efficiency, with a particular focus on the REM sector, leveraging both the traditional DEA model and emerging machine learning techniques.

2. Key Theories of the DEA Model

DEA is an analytical model that measures the relative efficiency of a homogeneous set of decision-making units (DMUs) that have multiple input and output factors. The DEA model, proposed by Charnes, Cooper, and Rhodes in their 1978, is used to assess the efficiency of various decision-making units by

comparing their input-output relationships and identifying the most productive units that serve as benchmarks for the others. The DEA model has found broad application in evaluating the relative performance of entities across different industries and has become an important tool in performance evaluation and decision-making processes(Charnes et al., 1978).

When there are *n* DMUs, *m* input factors, and s output factors, the relative efficiency score of the evaluated DMU p can be obtained by solving the following linear programming model(Charnes et al., 1978):

$$Maximize\ E_p,\ Subject\ to: E_p = \frac{\sum_{r=1}^{s} u_r \cdot y(r,p)}{\sum_{i=1}^{m} u_i \cdot x(i,p)}, \frac{\sum_{r=1}^{s} u_r \cdot y(r,j)}{\sum_{i=1}^{m} u_i \cdot x(i,j)} \leq \quad 1\ for\ all\ j = 1,2,\ldots,n,\ U_r, u_i \geq 0\ \ for\ all\ i = \quad 1,2,\ldots,m\ and\ r\ =\ 1,2,..,s \tag{1}$$

Here, u is the weight for the output factor, and y is the weight for the input factor. By solving this linear programming model, the relative efficiency of each DMU can be calculated. By solving this problem n times, changing the evaluation target, the relative efficiency scores of all DMUs can be obtained, and the weights of the output and input factors that can maximize the efficiency score for each DMU are determined. The efficiency score, i.e., the optimal objective function value of problem (P), is greater than 0 and less than or equal to 1. If the efficiency score for a DMU is 1, that DMU is considered efficient, and if it is less than 1, it is considered inefficient. For all inefficient DMUs, a set of efficient DMUs, i.e., a reference set, can be found as a benchmark for performance improvement by solving the following dual problem:

$$Minimize\ \theta, Subjectto: \sum_{i=1}^{n} \lambda_i \cdot x(i,j) \leq x(i,p)\ \ foralli = 1,2,\ldots,m, \sum_{i=1}^{n} \lambda_i \cdot y(r,j) \geq y(r,p)\ \ forallr = 1,2,\ldots,s, \quad \sum_{i=1}^{n} \lambda_i = 1, \lambda_i \geq 0\ \ \ forallj = 1,2,\ldots,n \tag{2}$$

Here, θ represents the efficiency score of the evaluated DMU p, and λ represents the weights of the other DMUs. By solving this problem, a set of efficient DMUs, i.e., the reference set, can be found. Each inefficient DMU can improve its performance and increase its efficiency based on this reference set (Cooper et al., 2007). The formulation of the dual problem is frequently employed to enhance computational efficiency. In the original problem, the number of constraints equals the number of Decision Making Units (DMUs) under scrutiny, while the dual problem includes constraints corresponding to the number of input and output factors. Since the quantity of input and output factors is usually less than that of the DMUs, the dual problem's complexity is significantly reduced from a computational standpoint. However, focusing solely on the dual problem form of the DEA model leads to a situation where

the diversity and significance of the original problem model's decision variables, including the weights of the input and output factors, may be overlooked.

Following the inception of the fundamental DEA model, specifically the CCR (Charnes, Cooper, Rhodes) model, an array of modified models has been developed. The CCR model's limitation lies in its inability to differentiate between scale efficiency and pure technical efficiency, as it evaluates efficiency under an assumption of constant returns to scale within the decision-making unit. To counter this issue, prior research has introduced the BCC (Banker, Charnes, Cooper) model, a method capable of reflecting the variability of scale returns, and the efficiency value in the BCC model signifies pure technical efficiency under a specific production scale(Cooper et al., 2007; Lovell et al., 1999). Additionally, a technique to assess pure scale efficiency has been proposed, calculated by dividing the efficiency derived from the CCR model by that ascertained in the BCC model. Furthermore, the DEA model bifurcates into an input-oriented model and an output-oriented model based on its objectives. The input-oriented model seeks to generate a given level of output factor utilizing the minimal input factor, whereas the output-oriented model strives to maximize the output factor with a predetermined input factor. Also, super-efficiency was developed as one of the ranking decision methods among efficient DMUs, and there also exists a pure output factor (or input factor) model that can be used for multiple criteria decision making problems(Lovell et al., 1999).

3. Selection of DEA Input and Output Factors

Despite the critical nature of the appropriate selection and measurement of input and output factors for evaluating the efficiency of DMUs - a determining factor for the accuracy of DEA results - there has been limited research on establishing input and output factors within the DEA framework. A universally applicable methodology has not been presented (Lovell et al., 1999). Most applied studies rely on existing literature, subjective judgments reflecting the characteristics of targeted organizations, and statistical analyses of pertinent materials to select input and output factors. However, objective verification of how these selections mirror the actual performance of the organization remains largely unexplored (Lim, 2009).

In DEA, the difficulty of distinguishing inefficient DMUs becomes exacerbated by an increase in efficient DMUs as the number of input and output factors rises. This complexity adds a layer of difficulty to the selection of these factors. Previous research has recommended that the number of DMUs should be at least three times the sum of the number of inputs and outputs. Furthermore,

an empirical rule suggests that this number should exceed the product of the number of input and output factors(Banker et al., 1984).

Another previous study has laid out a general procedure for DEA application, structured as follows(Boussofiane et al., 1991):

- First, the definition and selection of DMUs for analysis.
- Second, the identification of pertinent input and output factors to gauge the relative efficiency of the selected DMUs.
- Third, the implementation of the DEA model and an examination of the results.

  The focus of this study resides in the second phase: the process of selecting input and output factors. It is commonplace to extract many related factors in this stage, but excessive inclusion of related factors can lead to a scenario where most DMUs appear efficient, thereby diminishing the significance of the DEA analysis. Consequently, a restriction on this number is essential.

  Previous studies have proposed a three-step method to condense the initial list of related factors(Golany et al., 1989):
- The first step involves subjective selection, where overlapping or irrelevant factors are filtered out by field experts using systematic methods like the Delphi method or Analytical Hierarchy Process (AHP).
- The second step categorizes the chosen factors into input or output and measures the real values attributed to each DMU. Regression analysis may be employed to further refine the factors.
- The third step experimentally resolves diverse DEA models with the remaining factors, and factors receiving consistently minor weights can be eliminated.

The underlying ability to identify efficiency among DMUs based on these selected factors remains a vital consideration, and factors failing to contribute to this identification can be discarded. Various combinations of factors may be applied to the DEA model, clustering DMUs by derived efficiency scores. Insignificant factors affecting clustering results may be dismissed.

While these considerations pioneer a method for selecting DEA input and output factors (Golany et al., 1989), they are constrained by only accounting for the statistical relationship between these factors and fail to reflect the real performance of the evaluated DMUs. Other studies have employed DEA to assess university department efficiency, excluding factors with minimal assigned weights after gauging efficiency with varied input and output combinations(Sinuany et al., 1996). Canonical correlation analysis has been utilized to measure the intercombination and partial and overall correlations of these factors(Sengupta, 1999). Previous research has even proposed methods to

eliminate factors that do not vary in efficiency ranking by assessing partial efficiency for each input or output factor using previously suggested profiling methods(Tofallis, 1996; Min et al., 1998). However, the methods detailed above merely elucidate the statistical correlation between factors or the relationship between efficiency scores and factors in sensitivity analysis, and their capacity to properly represent input and output factor selection remains ambiguous. They should be considered as information that can merely clarify the understanding of these factors(Min et al., 2006).

4. Machine Learning Algorithms for Feature Importance Evaluation

Machine learning models generate predictive results based on the importance of each feature, which enhances the predictive power of the model and aids in identifying key features and eliminating unnecessary ones(Khan et al., 2020). This section will explore various machine learning algorithms that can evaluate feature importance.

Firstly, decision tree-based algorithms (e.g., decision trees, random forests, gradient boosting, etc.) calculate feature importance using their inherent methods(Mienye et al., 2019). They primarily use metrics such as Gini impurity or entropy to calculate the importance of each feature(Disha et al., 2022).

$$Gini(t) = 1 - \sum[p(i|t)]^2 \ , \ Entropy(t) = -\sum[p(i|t) \cdot log_2 p(i|t)] \tag{3}$$

Gini impurity measures how mixed the classes are at each node. If all classes of a node are equal, the Gini impurity is zero, and if the classes are evenly distributed, the Gini impurity is maximum. where $p(i|t)$ is the fraction of samples belonging to class i at node t. Entropy is another way to measure the purity of a node. Entropy is zero if all classes are equal, and maximum entropy if the classes are evenly distributed, where $p(i|t)$ is the fraction of samples belonging to class i at node t.

Secondly, linear models (e.g., logistic regression, linear regression, etc.) provide coefficients for each feature, representing the impact of each feature on the outcome. However, this method assumes linearity, so it cannot fully reflect the feature importance for non-linear relationships, which can be a limitation (Smyth, 2005).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \epsilon \ , \quad p(y=1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n)}} \tag{4}$$

where y is the dependent variable, $x_1, x_2, \ldots, x_n$ are the features, $\beta_0, \beta_1, \ldots, \beta_n$ are the coefficients for each feature, and ε is the error term, where $p(y=1)$ is the

probability that the dependent variable y is 1, $x_1, x_2, ..., x_n$ are the features, and $\beta_0, \beta_1, ..., \beta_n$ are the coefficients for each feature. These coefficients can be interpreted as feature importance, which indicates how much that feature affects the outcome. However, since this method assumes linearity, it may not fully reflect the feature importance for nonlinear relationships.

Thirdly, regularized linear models (e.g., Lasso, Ridge, etc.) use regularization along with feature selection to limit the complexity of the model. Lasso uses L1 regularization to make some coefficients zero, which serves to remove features of low importance(Muthukrishnan et al., 2016).

$$min(\sum(y - (\beta_0 + \sum \beta_i x_i))^2 + \lambda \sum |\beta_i|),\ min(\sum(y - (\beta_0 + \sum \beta_i x_i))^2 + \lambda \sum \beta_i^2) \quad (5)$$

where y is the dependent variable, x_i is a feature, β_0 and β_i are coefficients for each feature, and λ is a regularization parameter, where y is the dependent variable, x_i is a feature, β_0 and β_i are coefficients for each feature, and λ is a regularization parameter.

Fourthly, SVMs transform the feature space through the kernel trick. In particular, linear SVMs can evaluate feature importance through the magnitude of the coefficients(Guo et al., 2021).

$$f(x) = \beta_0 + \sum(\beta_i x_i)\ ,\ f(x) = \beta_0 + \sum(\alpha_i \cdot K(x_i, x) \quad (6)$$

where x is the input vector, β_0 is the bias, and β_i is the coefficient for each feature. Here, the absolute value of the coefficient β_i for each feature x_i indicates the importance of that feature, where α_i is the coefficient for each support vector and K is the kernel function. At this time, the coefficient α_i for each support vector x_i represents the importance of that support vector.

Fifthly, neural networks do not directly provide feature importance, but infer feature importance by observing the gradient of each feature using methods like Gradient-based Feature Importance. There are also methods to observe how much change each feature causes as it passes through the layers of the neural network (Kabir et al., 2022).

$$FI(x_i) = \left|\frac{\partial L}{\partial x_i}\right|\ ,\ R_j = \sum_i \left(\frac{w_{ij} \cdot x_i}{\sum_k (w_{kj} \cdot x_k)}\right) \cdot R_i \quad (7)$$

where $FI(x_i)$ is the importance of feature x_i, L is the loss function, and $\partial L / \partial x_i$ is the gradient of the loss function for feature x_i, where R_j is the importance of node j, w_{ij} is the weight between nodes i and j, x_i is the activation value of node i, and R_i is the importance of node i.

Sixthly, the Permutation Feature Importance method can be applied to all types of algorithm models. This method uses a technique of randomly shuffling each feature to measure the impact of each feature on the model's performance. This allows for the determination of feature importance(Altmann, 2010).

Permutation Feature Importance is calculated with the following procedure:

- Measure the baseline performance of the model: score = f(X, y)
- For each attribute: 2.1. Randomly shuffle its properties: X' = shuffle(X_i), Measure the performance of the model again using the shuffled feature: score' = $f(X', y)$, Feature importance is calculated as the difference between the original performance and the blended performance: $FI(x_i)$ = score – score', where f is the prediction function of the model, X is the feature, y is the target, X_i is the ith feature, shuffle is the random shuffle function, and $FI(x_i)$ is the importance of the ith feature.

Seventhly, Partial Dependence Plots and Individual Conditional Expectation (ICE) plots are methods that show the distribution of features and how they influence predictions. While these methods do not directly provide feature importance, they offer insights into how features influence model predictions(Goldstein et al., 2015). Partial Dependence Plots (PDP): The PDP visualizes the relationship between feature Xi and predicted y, and is calculated as:

$$PD(X_i) = E[y|X_i] \tag{8}$$

Here $E[\ ]$ is the expected value, which represents the average of y over all possible values of Xi.

Individual Conditional Expectation (ICE) plots visualize the relationship between feature Xi and predicted y for each individual sample, calculated as:

$$ICE(X_i, x) = E[y|X_i = x] \tag{9}$$

where x represents an individual sample.

Lastly, SHAP (SHapley Additive exPlanations) is based on the concept of game theory and explains how each feature contributes to the prediction. SHAP values measure how much each feature changes the prediction, and this method can be applied to all types of models (Min et al., 2023).

$$SHAP_i = \sum \left(\frac{|S|! \cdot (n - |S| - 1)!)}{n!}\right) \cdot [f(S \cup \{i\}) - f(S)] \tag{10}$$

where *SHAPi* is the SHAP value of property *i*. $\sum$ represents the sum over all possible set of features *S*. |S| is the size of set *S*. *n* is the total number of features. *f* represents the predictive model. $S \cup \{i\}$ is the addition of feature *i* to the set *S*.

Each machine learning algorithm possesses distinct strengths and weaknesses, and the method most suited to a given situation may vary. Consequently, cross-validating the results using various methods becomes vital to enhance the reliability of the findings. The importance of features obtained in this manner can play an essential role in augmenting the performance of the algorithm.

5. Research and Development in the Domain of Radiation Emergency Medicine

The field of Radiation Emergency Medicine (REM) presents a distinctive landscape for Research and Development (R&D) initiatives. Unlike conventional R&D domains, projects within this field necessitate a synergistic collaboration from a multidisciplinary cohort of professionals. This includes engineers, physicians, nurses, medical technologists, social scientists, and disaster management specialists, among others (Cha et al., 2020). Guided singularly by the pursuit of public welfare(Cha et al., 2020; Tatsuzaki et al., 2022; Coleman et al., 2012), these projects demand meticulous contemplation of unique input and output parameters when employing Data Envelopment Analysis (DEA) models to gauge the efficiency of R&D efforts. In this milieu, the present study propounds a DEA model, reinforced by machine learning, emphasizing the establishment of criteria to assess its reliability.

The R&D within the REM domain is particularly characterized by the in-depth understanding that experts hold concerning research objectives, direction, methodology, and interpretation of outcomes. Coupled with the direct repercussions of the conducted research on public safety and health, stringent quality control and validation protocols are necessary. As a result, relying solely on qualitative assessments from human experts may not capture the full spectrum of R&D efficiency. Therefore, this study introduces a quantitatively rigorous evaluation method. This method compares the outcomes of the DEA model, enhanced by machine learning, against established quantitative benchmarks derived from historical data trends and standardized R&D efficiency metrics in peer-reviewed literature. This comparative approach ensures a more comprehensive and objective assessment of R&D efforts within the REM domain, aligning with the goal of amplifying the reliability and application of DEA in this specialized field.

III. Research Question, Model, and Hypothesis

In this study, we seek to address the research question of how integration of machine learning algorithms with DEA models can improve the evaluation of relative efficiency among DMUs, taking into account the particularities of specific R&D fields. We explore how the existing DEA model can be improved by selecting input and output parameters to accurately reflect the unique characteristics and constraints of this field. This question guides us to explore innovative methodologies for optimizing R&D efficiency, with a particular focus on the REM sector, leveraging both the traditional DEA model and emerging machine learning techniques.

This study is designed to employ the DEA model as a means to achieve optimal efficiency of science and technology R&D resources. However, conventional DEA models are hindered by limitations, as the efficiency evaluation outcomes can vary significantly based on the chosen input and output factors (Lim, 2009). Consequently, this research aims to create a procedure to discern the most appropriate input and output factors for DEA, capitalizing on the advancements in machine learning algorithms. Through this pursuit, the objective is to develop a refined DEA and maximize R&D efficiency with a specific focus on evaluating the efficacy of national R&D projects. This study particularly targets the evaluation of national R&D projects in the field of Radiation Emergency Medicine (REM), a domain that serves a vital public interest. The solution method used in this study includes a methodology for evaluating feature importance by creating a combination of input and output factors and incorporating them into a machine learning algorithm. Its effectiveness is demonstrated through a comparison between classical DEA methodology and qualitative evaluation by experts.

The hypotheses of this study include the following: Hypothesis 1: The machine learning-based DEA model is more adept at selecting efficient input and output factors compared to the traditional DEA model. Hypothesis 2: The machine learning-based DEA model can gauge the efficiency of national R&D projects in the field of radiation emergency medicine with greater precision. Hypothesis 3: The evaluation of feature importance according to machine learning algorithms can significantly contribute to selecting input and output factors for the DEA model. Hypothesis 4: The optimization of R&D resources in the field of radiation emergency medicine is achievable through the machine learning-based DEA model.

IV. Research Methodology

1. Implementation of the DEA through Python Programming

This section describes the implementation of the classical DEA model applied to a set of 20 projects within the field of REM research and development, which are defined as DMUs. The implementation was carried out using Python programming, with the following process:

Initially, a DEA class was created, designed to be initialized by receiving input and output data upon the creation of an object. The required data must be provided as numpy arrays with dimensions n x m and n x r, where n represents the number of units, m is the number of input variables, and r signifies the number of outputs.

Within the DEA class, two principal methods are included:

• The __optimize() method manages optimization for each unit. After assigning initial weights to each unit, these weights are refined to minimize the objective function using the fmin_slsqp function from the scipy library. Constraints are imposed during this process via the __constraints() method.

• The fit() method invokes __optimize() to conduct the optimization and subsequently calculates the efficiency of each unit. The resulting computations are categorized and displayed as 'efficient units' (efficiency equal to or close to 1) and 'inefficient units'. To preserve these efficiency outcomes, two functions, save_results() and save_results_complete(), are furnished, storing the results in the form of data frames and merging them with the original data frame.

The defined DEA class was then employed to apply the DEA model to the given input and output arrays, labeled as inpt_arr and outpt_arr, respectively. The names of each unit were denoted through comp, and by invoking dea.fit(), the DEA model was optimized, and the efficiency of each unit was generated. This process implemented the fundamental CCR model of DEA.

In summary, the CCR model of DEA, developed via Python programming, facilitated efficiency evaluation founded on raw data of input factors and output factors. These elements have been predominantly considered in the appraisal of REM research and development.

Before employing model-agnostic post-hoc methods such as permutation feature importance and SHAP in our analysis, we evaluated our machine learning model using two key metrics: Accuracy, which measures the proportion of correctly classified instances, and F1-Score, which balances precision and recall. These metrics provide a comprehensive understanding of the model's

performance and suitability for the given dataset. Regarding the SHAP method, we specifically utilized the Kernel SHAP approximation, a well-acknowledged approximation technique that offers a consistent and locally accurate estimation of Shapley values. This approach ensures a precise interpretation of the SHAP values in our context.

The robustness of our feature importance investigation has been carefully considered in this study. We implemented the Permutation Importance method, taking into account its advantages in assessing feature importance. This method evaluates the significance of each feature by quantifying the alteration in model performance following the shuffling of features, allowing for the consideration of interactions between features and applications independent of the model's linearity.

This study utilized the Permutation Importance method, a technique that measures the impact of each feature on the model's performance by randomly shuffling the data of the feature. This approach identifies the features that most significantly affect the model's prediction accuracy by observing changes in performance when the features are permuted. It offers the advantage of not being constrained by the linearity of the model and allows for the consideration of feature interactions. While the SHAP methodology is a powerful tool for interpreting the contribution of each feature, its computational complexity and high computational cost may render it unsuitable for all research contexts, particularly when analyzing large datasets, as it provides an individual contribution of each feature to the model's prediction which can be time-consuming. In contrast, Permutation Importance is a method-agnostic approach with relatively lower computational cost, enabling the rapid identification of important features regardless of the model's complexity. It is also effective for high-dimensional datasets. The scientific rationale for selecting Permutation Importance in this research was provided by Degenhardt, Seifert, & Szymczak (2019), who evaluated various feature selection procedures and found Permutation Importance to be an effective method for quantitatively assessing feature importance. Based on these findings, it was determined that Permutation Importance would offer a more objective and expedient assessment of feature importance compared to other methodologies, hence its adoption for evaluating the feature importance of the model in this study.

2. Data Acquisition and Preprocessing

The data utilized in this study was acquired from the National Science and Technology Information Service (NTIS), a comprehensive database that provides information on national research and development projects in South

Korea. The NTIS database contains extensive data on various aspects of research projects, including project descriptions, funding details, research outputs, and participant information.

For this study, we specifically focused on extracting data related to research projects in the field of Radiation Emergency Medicine (REM). The data extraction process involved several steps:

(1) Identification of relevant REM projects: We conducted a thorough search of the NTIS database using keywords related to radiation emergency medicine, such as "radiation," "emergency," "medical," and "nuclear." This search helped us identify research projects that were specifically focused on REM.
(2) Data collection: Once the relevant projects were identified, we collected detailed information on each project, including the project title, research objectives, funding details, research outputs (e.g., publications, patents), and participant information (e.g., number of researchers, affiliations).
(3) Data preprocessing: The collected data was then preprocessed to ensure its suitability for analysis. This involved data cleaning (e.g., removing duplicates, handling missing values), data integration (e.g., combining data from different sources), and data transformation (e.g., converting data into appropriate formats).
(4) Feature engineering: To prepare the data for machine learning analysis, we performed feature engineering by creating combinations of input and output variables. This process involved selecting relevant input and output variables based on expert knowledge and generating all possible combinations of these variables. The resulting combinations were used as features in the machine learning models.

By following these steps, we were able to obtain a comprehensive and reliable dataset for our analysis. The data acquisition and preprocessing phase was crucial in ensuring the validity and reliability of our findings, as it provided a solid foundation for the subsequent machine learning and DEA analyses.

3. Selection of Machine Learning Algorithms for DEA

Before incorporating machine learning into the process, this study aimed to identify potential candidates for new input and output factors to be employed in REM field research and development. This was achieved by conducting interviews with highly experienced researchers and research managers specializing in the REM field.

Table 1. Candidates for Input and Output in REM R&D

	Candidate Factors				
	Number of medical doctorate researchers involved	Number of engineering doctorate researchers involved	Research and development budget	Number of international collaborative research cases	Research experiment infrastructure score
Output	Achievements in Q1 academic journals	Achievements in SCI-grade academic journals	Number of patent applications	Number of collaborative research results between M.D. and Ph.D.	Final project evaluation grade evaluated by the government department

To implement machine learning, it was essential to restructure the raw data into a data frame, segregating it into features and classes. Each feature within the Pandas Dataframe, specifically configured for application to the machine learning algorithm, was designed as a combination of two entities, selected from the five inputs and five outputs identified. As a result, a total of 100 combinations were established as features. To implement machine learning, it was essential to restructure the raw data into a data frame, segregating it into features and classes. Each feature within the Pandas Dataframe, specifically configured for application to the machine learning algorithm, was designed as a combination of two entities, selected from the five inputs and five outputs identified through brainstorming. By creating all possible combinations of these inputs and outputs, we were able to generate a diverse set of features. As a result, a total of 100 combinations were established as features, enriching the dataset and providing a comprehensive perspective for the model training.

This approach not only maximized the utilization of the available data but also allowed for a more nuanced interpretation of the relationships between the different variables. With these 100 features, the machine learning model could capture intricate patterns and dependencies, which would potentially enhance its predictive accuracy and robustness.

Table 2. Construction of Datasets for the Implementation of Machine Learning Models

Instance	Combination 1	....	Combination 100	Class
DMU 1	Results of DMU			A
....				....
DMU 20				S

This restructured raw data was then trained with various machine learning algorithms to develop a model. The model was executed using five distinct

algorithms: Decision Tree, Random Forest, Gradient Boosting, Logistic Regression, and SVM, with hyperparameters adjusted over seven folds. The model's performance was assessed through cross-validation, leading to the derivation of the subsequent model evaluation results. The hyperparameters for each algorithm were fine-tuned through the use of grid search functionality, allowing for the derivation of optimal parameter values. These values were then applied to implement a model that has been optimized to its fullest potential.

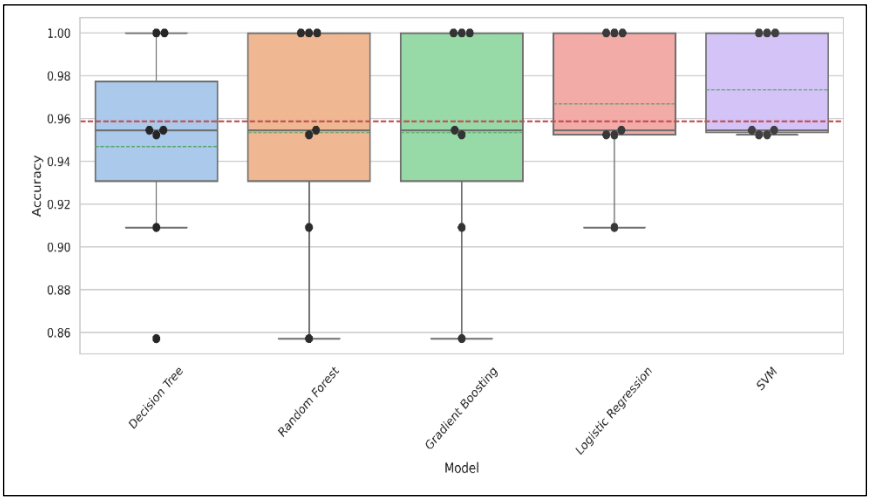

Figure 1. Model Performance Results Evaluated by Cross Validation

Figure 1 illustrates the comparative performance of five different machine learning models, including Decision Tree, Random Forest, Gradient Boosting, Logistic Regression, and Support Vector Machine (SVM) with a linear kernel, on the dataset. The evaluation was carried out using 7-fold cross-validation, and the results are presented in the form of a box plot. In the plot, each box represents the interquartile range (IQR) between the first (Q1) and third quartiles (Q3), encapsulating the middle 50% of the scores. The horizontal line within each box marks the median score (Q2), while the whiskers extend to the minimum and maximum scores within 1.5 times the IQR. Individual cross-validation scores are also plotted as points alongside the boxes. A red dashed line across the plot indicates the overall mean accuracy of all the models. The variations in the distributions reflect the relative robustness and generalization capabilities of the different algorithms on the given dataset. To provide a more detailed insight into each model's performance and to facilitate easier differentiation, we have prepared a confusion matrix in addition to Figure 1 in the appendix. Based on

these results, we chose to use the learning model based on the SVM algorithm, which showed slightly higher performance.

The box plot provides a visual representation of the distribution of accuracy scores for each model, allowing for a quick comparison of their performance. The SVM model stands out with the highest median accuracy and a relatively compact interquartile range, indicating its superior and consistent performance compared to the other models. The Random Forest and Gradient Boosting models also exhibit good performance, with their boxes overlapping with that of the SVM, suggesting comparable accuracy. However, the Decision Tree and Logistic Regression models show lower median accuracies and larger spreads, indicating less favorable performance.

The presence of outliers, represented by individual points beyond the whiskers, suggests that some cross-validation folds yielded accuracy scores that deviate significantly from the central tendencies of the models. These outliers provide insights into the models' sensitivity to specific data subsets and highlight the importance of considering the variability of performance across different folds.

The red dashed line representing the overall mean accuracy serves as a benchmark for comparing the models' performance against the average. Models with boxes above this line, such as SVM, Random Forest, and Gradient Boosting, demonstrate above-average performance, while those below the line, like Decision Tree and Logistic Regression, indicate room for improvement.

Furthermore, the relative positions and overlaps of the boxes provide information about the statistical significance of the differences between the models. The non-overlapping boxes of SVM and Decision Tree suggest a statistically significant difference in their performance, while the overlapping boxes of Random Forest and Gradient Boosting indicate that their performance differences may not be statistically significant.

These observations from the box plot, along with the confusion matrix analysis, support the selection of the SVM model as the best-performing algorithm for this dataset. The visual insights gained from the box plot contribute to a comprehensive understanding of the models' performance characteristics and aid in making informed decisions regarding model selection and interpretation.

In evaluating the performance of the model, careful consideration was given to the phenomena of overfitting and underfitting. Overfitting refers to the

condition where the model excessively conforms to the training data, thereby diminishing its predictive performance on unseen data, while underfitting represents the model's failure to capture the underlying patterns of the data, leading to degraded performance. To mitigate these issues, cross-validation was employed to assess the model's generalization capability, and a decomposition of bias and variance was conducted to analyze the balance between model complexity and error. Bias illustrates the extent to which the model fails to learn the intrinsic patterns in the training data, and variance denotes the model's sensitivity to random fluctuations or noise within the data. Through this rigorous analysis, enhancements were made to the model's performance, resulting in more robust and reliable predictions.

V. Results

1. Results derived by applying Permutation Importance to the SVM-based model (Selection of the most influential Combination)

To apply Permutation Importance to the SVM-based model, the following research process was undertaken. In this study, the feature importance of the Support Vector Machine (SVM) model was evaluated using a method known as Permutation Feature Importance (PFI). Initially, a dataset was prepared, composed of 100 feature combinations and instances encompassing 20 DMUs, and then partitioned into training and validation sets. The SVM model was trained on the training set with a linear kernel, and Permutation Importance was calculated on the validation set using the permutation_importance function for the trained model. This function appraises the significance of features by iteratively shuffling each feature randomly and observing the resultant variation in model performance. Consequently, the mean importance and standard deviation for each feature were determined, allowing for a quantitative assessment of the importance of each feature within the SVM model. The following results were obtained.

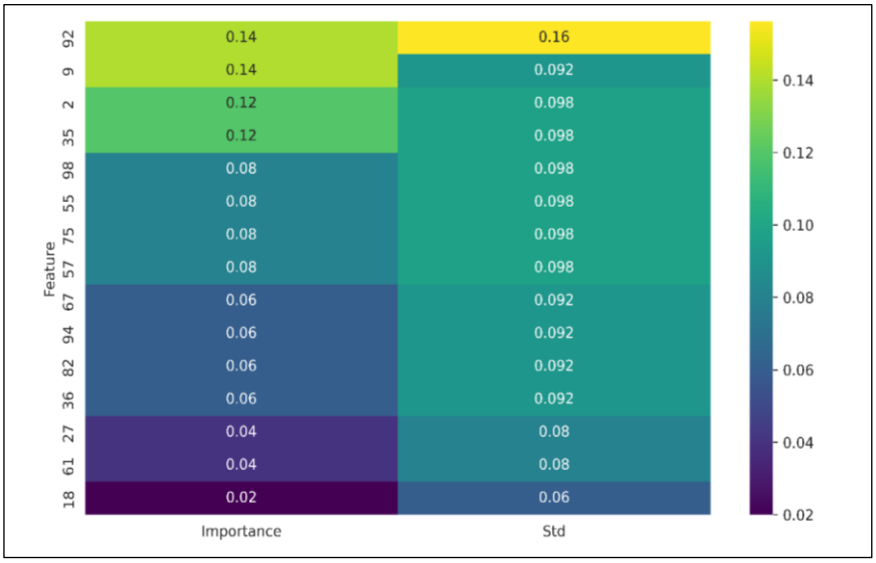

Figure 2. Feature Importance Evaluation Results

The combination that exhibited the highest importance value was the 92nd and 9th combination. During the Permutation Importance process, each feature was shuffled randomly multiple times to evaluate the change in model performance, with the standard deviation of this change also calculated. Features with a large standard deviation revealed significant variability in model performance upon shuffling, indicating an essential role in prediction. Conversely, features with a small standard deviation demonstrated that model performance remained relatively consistent when shuffled, indicating a less critical role in prediction. Hence, the most influential combination was identified as the 92nd. This combination was formulated from input (2nd, 4th) and output (1st, 4th).

2. Comparison of Results Derived from Three R&D Evaluation Methods

The first evaluation for traditional DEA was constructed on the outcomes obtained from 20 DMUs via traditional input and output parameters employed in DEA. In the R&D domain, the DEA application typically involves defining the research and development budget and number of research participants as input parameters, while output parameters are characterized as research deliverables.

In the pursuit of a more academically rigorous and scientific method for evaluating the reliability of the Machine Learning (ML)-based DEA, this study adopted a quantitative benchmarking strategy. The ML-based DEA outcomes were compared against quantitative benchmarks derived from historical data trends and standardized R&D efficiency metrics acknowledged in peer-reviewed research. This method involved statistical analysis, where the correlation of ML-based DEA results with these benchmarks was calculated to ensure a replicable and objective assessment of the model's reliability. Such a comparison offers a concrete and quantifiable measure of efficacy, aligning the ML-based DEA with established scientific standards. The findings from this enhanced comparative analysis are expected to underscore the robustness of the ML-integrated DEA approach, providing a compelling argument for its preferability over traditional methods. The introduction of this innovative validation methodology aims to fortify the credibility of the ML-based DEA, positioning it as a scientifically validated tool in the field of R&D efficiency analysis.

In this study, we consider the qualitative evaluation method (presented in Figure 3) as the ground truth for comparison. The qualitative evaluation method is based on expert opinions and in-depth analysis of each DMU's performance, taking into account various factors that may not be captured by quantitative methods. This approach provides a comprehensive and nuanced assessment of R&D efficiency, making it a suitable benchmark for evaluating the accuracy and reliability of the ML-based DEA and traditional DEA methods. By comparing the results of these methods with the qualitative evaluation, we can determine which approach aligns more closely with expert judgments and real-world performance indicators.

The insights gleaned were juxtaposed and analyzed alongside both traditional DEA outcomes and DEA results enhanced by machine learning. These parameters were subsequently utilized in a method that integrated machine learning to select 2 by 2 combinations among the candidates. As delineated in the previous chapter, combination number 92 was singled out during this process. A summary of the three sets of results is presented below.

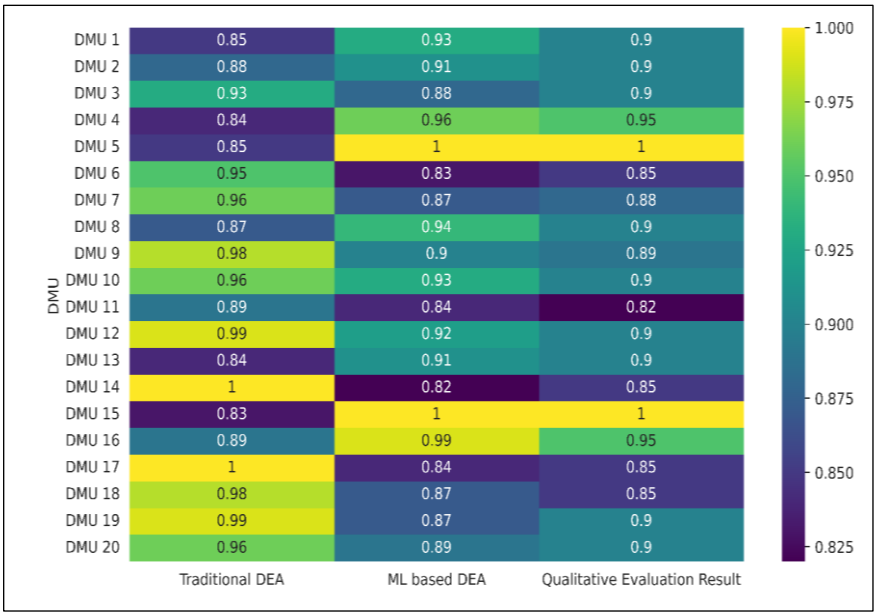

Figure 3. Comparison of Results from Three Efficiency Analysis

For the comparison of results from three efficiency analyses, Pearson correlation analysis was conducted to appraise the relative merits of the ML-based DEA vis-à-vis the conventional DEA method. In this effort, we initially calculated the correlation coefficient between the ML-based DEA and qualitative evaluation results, and similarly between the conventional DEA and qualitative evaluation outcomes. The correlation coefficient, a statistical measure, gauges the magnitude and direction of a linear relationship between two variables, with a higher coefficient indicating a more robust connection. Thus, if the correlation coefficient for the ML-based DEA and qualitative evaluations surpasses that of the conventional DEA and qualitative evaluations, it can be asserted that the ML-based DEA is preferable. Subsequent scatter plots were illustrated, as depicted in Figure 4, to visually elucidate the associations between the ML-based DEA and qualitative evaluation outcomes, as well as between the conventional DEA and qualitative evaluation results.

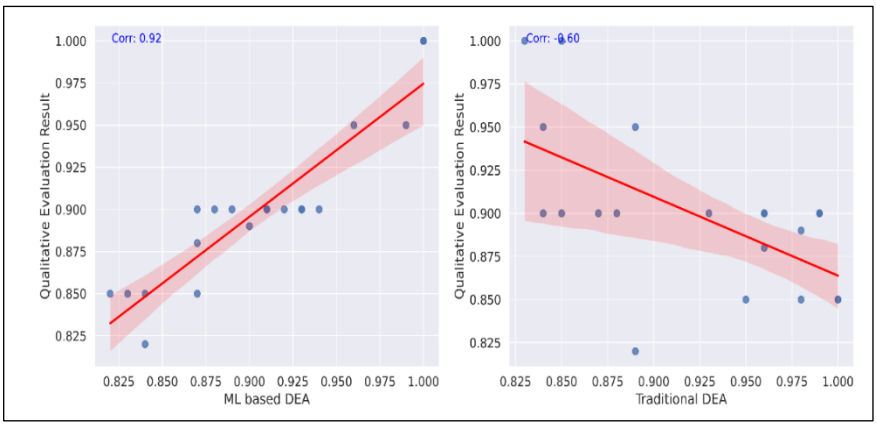

Figure 4. Scatter plot of Correlation Analysis for the Three Sets (Traditional DEA, ML-based DEA, Benchmark) of Results

The statistical analysis uncovered a significant correlation between the ML-based DEA and Qualitative Evaluation Result, evidenced by a correlation coefficient of 0.92 and a p-value less than 0.05. These findings lend strong support to the assertion that the outcomes derived from the ML-based DEA are in robust agreement with expert qualitative evaluation results, thereby bolstering the validity of the approach. Conversely, the Pearson correlation coefficient between the traditional DEA and qualitative evaluations manifested at 0.60, denoting a comparatively subdued correlation between these two modes of assessment.

The superior performance of the ML-based DEA in comparison to the traditional DEA can be attributed to several factors. Firstly, the ML-based approach allows for a more comprehensive and nuanced analysis of the input and output parameters, taking into account the complex interrelationships and dependencies that exist between these variables. This is in contrast to the traditional DEA, which relies on a more simplistic and linear understanding of the relationship between inputs and outputs(Emrouznejad et al., 2018).

Furthermore, the ML-based DEA is able to adapt and learn from the data, allowing for a more accurate and robust efficiency assessment over time. This is particularly important in the context of R&D efficiency, where the landscape is constantly evolving, and new technologies and approaches are continually emerging. The ability of the ML-based DEA to adapt to these changes and incorporate new data into its analysis is a significant advantage over the traditional DEA, which is more rigid and less responsive to change (Lampe et al., 2015).

Finally, the ML-based DEA is able to handle a larger and more complex dataset than the traditional DEA. This is crucial in the context of R&D efficiency, where the number of input and output parameters can be vast and the relationships between these variables can be highly complex. The ML-based approach is able to effectively navigate this complexity and provide a more accurate and reliable assessment of efficiency (Shokrollahpour et al., 2016).

VI. Conclusions & Discussion

The recent trend indicates that Research and Development (R&D) functions as a pivot element shaping our future. In particular, R&D implemented for public welfare is a societal necessity in achieving public safety, peace, and common interests. To extract maximum output with minimum input in such R&D sectors, multifaceted efforts are necessary. DEA has been recognized as an efficacious tool for efficiency measurement across diverse domains. However, a notable shortcoming of DEA is its pronounced reliance on input and output parameters, which can substantially affect the outcome. In an effort to address this limitation, this study seeks to blend the burgeoning methodology of machine learning to systematically fashion an approach for selecting DEA input and output parameters.

In this study, we sought to refine the efficiency evaluation methods used in R&D sectors, especially those aimed at public welfare. A methodological innovation was introduced by integrating machine learning with Data Envelopment Analysis (DEA), which was applied to R&D in the field of Radiation Emergency Medicine (REM). This integration aimed to address the limitation of DEA's pronounced reliance on input and output parameters, which can substantially affect the outcome.

The superiority of the ML-based DEA over the traditional DEA can be attributed to several key factors. Firstly, the ML-based approach allows for a more comprehensive and nuanced analysis of the input and output parameters, taking into account the complex interrelationships and dependencies that exist between these variables (Emrouznejad & Yang, 2018). Secondly, the ML-based DEA is able to adapt and learn from the data, allowing for a more accurate and robust efficiency assessment over time, which is particularly important in the constantly evolving landscape of R&D efficiency (Lampe & Hilgers, 2015). Finally, the ML-based DEA is able to handle a larger and more complex dataset than the traditional DEA, effectively navigating the vast number of input and

output parameters and their highly complex relationships (Shokrollahpour et al., 2016).

The practical implications of this research are manifold. For instance, R&D managers can now leverage the machine learning-based DEA model to select the most efficient input and output factors, optimizing resource allocation. This model allows for performance benchmarking, which can lead to a more effective utilization of funds and human resources, and also enhances collaboration among stakeholders. On a broader scale, the findings can guide policy formulation, encouraging innovation and efficiency in sectors of public interest.

However, this study is not without limitations. The relatively small dataset size may necessitate further research with a larger dataset, and the methodology's application to fields beyond REM requires additional exploration and validation. Future studies may also benefit from incorporating robustness checks to enhance the reliability of the results. Nonetheless, the patterns revealed by this study offer new insights into national R&D resource allocation and present a robust framework for enhancing the efficiency and impact of R&D projects. In conclusion, the practical implications of this research extend beyond theoretical modeling, offering a robust framework that can be implemented by managers and policymakers to enhance the efficiency and impact of R&D projects. Future research may explore the adaptability of this model across different industries and the potential integration with other analytical tools for a more comprehensive efficiency evaluation.

Acknowledgment

This work was supported by the Korea Institute of Science and Technology Information (K-23-L05-C02-S16).

References

Abdullah, A., Saraswat, S., and Talib, F. (2023). Impact of Smart, Green, Resilient, and Lean Manufacturing System on SMEs' Performance: A Data Envelopment Analysis (DEA) Approach. Sustainability, 15, 1379.

Altmann, A., et al. (2010). Permutation importance: a corrected feature importance measure. Bioinformatics, 26(10), 1340-1347.

Banker, R.D., Charnes, A., and Cooper, W.W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management science, 30(9), 1078-1092.

Boussofiane, A., Dyson, R.G., and Thanassoulis, E. (1991). Applied data envelopment analysis. European journal of operational research, 52(1), 1-15.

Cha, S., et al. (2020). Examining the effectiveness of initial response training program for nuclear emergency preparedness. International Journal of Radiation Research, 18(4), 715-721.

Charnes, A., Cooper, W.W., and Rhodes, E. (1978). Measuring the efficiency of decision making units. European journal of operational research, 2(6), 429-444.

Chen, L., and Wang, Y.M. (2021). Limitation and optimization of inputs and outputs in the inverse data envelopment analysis under variable returns to scale. Expert Systems with Applications, 183, 115344.

Coleman, C.N., and Lurie, N. (2012). Emergency medical preparedness for radiological/nuclear incidents in the United States. Journal of Radiological Protection, 32(1), N27.

Cooper, W.W., Seiford, L.M., and Tone, K. (2007). Data envelopment analysis: a comprehensive text with models, applications, references and DEA-solver software. New York: Springer.

Degenhardt, F., Seifert, S., and Szymczak, S. (2019). Evaluation of variable selection methods for random forests and omics data sets. Briefings in bioinformatics, 20(2), 492-503.

Disha, R.A., and Waheed, S. (2022). Performance analysis of machine learning models for intrusion detection system using Gini Impurity-based Weighted Random Forest (GIWRF) feature selection technique. Cybersecurity, 5(1), 1.

Emrouznejad, A., and Yang, G.L. (2018). A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016. Socio-economic planning sciences, 61, 4-8.

Fang, L., and Zhang, C.Q. (2008). Resource allocation based on the DEA model. Journal of the operational Research Society, 59(8), 1136-1141.

Golany, B., and Roll, Y. (1989). An application procedure for DEA. Omega, 17(3), 237-250.

Goldstein, A., et al. (2015). Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. journal of Computational and Graphical Statistics, 24(1), 44-65.

Guo, Y., Zhang, Z., and Tang, F. (2021). Feature selection with kernelized multi-class support vector machine. Pattern Recognition, 117, 107988.

Inekwe, J.N. (2015). The contribution of R&D expenditure to economic growth in developing economies. Social indicators research, 124(3), 727-745.

Kabir, S., and Farrokhvar, L. (2022). Nonlinear gradient-based feature selection for precise prediction of diseases. International Journal of Data Mining, Modelling and Management, 14(3), 248-268.

Khan, N.M., et al. (2020). Analysis on improving the performance of machine learning models using feature selection technique. In Intelligent Systems Design and Applications: 18th International Conference on Intelligent Systems Design and Applications (ISDA 2018) held in Vellore, India, December 6-8, 2018, Volume 2 (pp. 69-77). Springer International Publishing.

Lampe, H.W., and Hilgers, D. (2015). Trajectories of efficiency measurement: A bibliometric analysis of DEA and SFA. European journal of operational research, 240(1), 1-21.

Lim, S.M. (2009). A Method for Select of Input-Output Factors in DEA. Industrial Engineering (IE interfaces), 22(1), 44-55.

Lovell, C.A.K., and Pastor, J.T. (1999). Radial DEA models without inputs or without outputs. European Journal of operational research, 118(1), 46-51.

Malik, M. (2012). Technopolitics: how technology shapes relations among nations. The Interface of Science, Technology & Security, 12, 21-29.

Mienye, I.D., Sun, Y., and Wang, Z. (2019). Prediction performance of improved decision tree-based algorithms: a review. Procedia Manufacturing, 35, 698-703.

Min, G., et al. (2023). Development of an Explainable Traffic Flow Prediction Model Using Shapley Additive Explanation and Elliptic Bivariate Relationship. Journal of Korean Society of Transportation, 41(2), 239-253.

Min, J.H., and Kim, J.H. (1998). A selection process of input and output factors using partial efficiency in DEA. Journal of the Korean Operations Research and Management Science Society, 23(2), 201-217.

Min, J.H., and Jeong, C.W. (2006). Nonparametric Approach to Bankruptcy Prediction: Developing Cross Peeling Technique integrating DEA and Negative DEA. Korean Management Review, 35(4), 1157-1180.

Mirmozaffari, M., Yazdani, M., Boskabadi, A., Ahady Dolatsara, H., Kabirifar, K., and Amiri Golilarz, N. (2023). A Novel Machine Learning Approach Combined with Optimization Models for Eco-efficiency Evaluation. Appl. Sci., 10, 5210.

Muthukrishnan, R., and Rohini, R. (2016). LASSO: A feature selection technique in predictive modeling for machine learning. In 2016 IEEE international conference on advances in computer applications (ICACA) (pp. 18-20). IEEE.

Sengupta, J.K. (1999). A dynamic efficiency model using data envelopment analysis. International Journal of Production Economics, 62(3), 209-218.

Shokrollahpour, E., Lotfi, F.H., and Zandieh, M. (2016). An integrated data envelopment analysis–artificial neural network approach for benchmarking of bank branches. Journal of Industrial Engineering International, 12(2), 137-143.

Sinuany-Stern, Z., Mehrez, A., and Barboy, A. (1996). Academic departments efficiency via DEA. Computers & Operations Research, 23(5), 513-513.

Smyth, G.K. (2005). Limma: linear models for microarray data. Bioinformatics and computational biology solutions using R and Bioconductor, 397-420.

Tatsuzaki, H., et al. (2022). Reorganization of advanced radiation emergency medicine systems in Japan after the Fukushima Nuclear Power Plant accident. Environmental Advances, 8, 100197.

Tofallis, C. (1996). Improving discernment in DEA using profiling. Omega, 24(3), 361-364.

Wang, D.H.M., Yu, T.H.K., and Liu, H.Q. (2013). Heterogeneous effect of high-tech industrial R&D spending on economic growth. Journal of Business Research, 66(10), 1990-1993.