Probability-based Efficiency Analysis through Machine Learning Techniques

Ricardo González-Moyano[1], Juan Aparicio[1,2*], Víctor J. España[1] and José L. Zofío[3,4]

[1] Center of Operations Research (CIO). Miguel Hernandez University, Elche, Spain.

[2] ValgrAI - Valencian Graduate School and Research Network of Artificial Intelligence, Valencia, Spain.

[3] Department of Economics, Universidad Autónoma de Madrid, Madrid, Spain.

[4] Erasmus Research Institute of Management, Erasmus University, Rotterdam, The Netherlands.

[*] Corresponding author: j.aparicio@umh.es.

Abstract

The integration of Data Envelopment Analysis (DEA) and Machine Learning (ML) has paved the way for more robust efficiency assessments of decision-making units (DMUs). Traditional DEA models, while effective, present challenges such as sensitivity to noise, deterministic outputs, and limited ability to account for uncertainty. This paper introduces an innovative classification-based ML approach to enhance DEA by incorporating probabilistic efficiency analysis. Our methodology reformulates DEA as a classification problem, using neural networks to predict efficiency status and estimate the probability of a DMU being classified as efficient. This enables counterfactual-based efficiency benchmarking, where the minimum adjustments required for an inefficient DMU to become efficient are systematically identified. A key contribution of this approach is its explainability, achieved through Explainable Artificial Intelligence (XAI) techniques, specifically sensitivity analysis (SA) and counterfactual explanations. We employ variable importance methods to provide targeted efficiency improvement strategies, ensuring that DMUs receive actionable recommendations based on their specific inefficiencies. Additionally, our framework introduces dynamic peer selection at different probability thresholds, allowing for adaptable benchmarking strategies. The proposed methodology is validated using a real-world dataset from Spain's food industry, demonstrating its applicability and reliability. Results highlight the advantages of this ML-DEA hybrid framework, including greater inferential power, improved discriminatory ability, and enhanced robustness in high-dimensional settings

Keywords: Data Envelopment Analysis, Machine Learning, Classification models, Neural Networks, variable importance.

1. Introduction

In recent decades, the field of efficiency analysis has witnessed significant advancements, particularly in the evaluation of decision-making units (DMUs) across various sectors such as finance, healthcare, education, and manufacturing. One prominent methodology that has garnered substantial attention is Data Envelopment Analysis (DEA), initially introduced by Charnes, Cooper, and Rhodes in the late 1970s (Charnes et al., 1978). DEA offers a non-parametric approach to assess the relative efficiency of DMUs by comparing their input-output profiles. The fundamental premise of DEA lies in its ability to evaluate the efficiency of DMUs that operate under multiple inputs and outputs, without imposing restrictive assumptions about functional forms or underlying distributions. This characteristic makes DEA particularly appealing for analysing complex real-world systems where the relationships between inputs and outputs use to be probably nonlinear and unknown. Over the years, DEA has been applied to diverse domains, including banking (Berger et al., 1997), healthcare (Olesen et al., 2007), and environmental performance assessment (Zhou et al., 2008), among others.

However, despite its widespread adoption and commendable performance, traditional DEA approaches may encounter limitations in capturing the intricate patterns and structures inherent in complex datasets. One notable challenge lies in the potential for overfitting, wherein the model captures noise or idiosyncratic features in the data rather than true underlying relationships (Esteve et al., 2020). This issue is particularly pronounced in DEA when dealing with high-dimensional datasets or when the number of DMUs is relatively small compared to the number of inputs and outputs, where overfitting is mixed with the curse of dimensionality problem (Charles et al., 2019). Then, DEA can lead to inflated efficiency scores for certain DMUs, thereby distorting the assessment of relative efficiency and potentially misleading decision-makers. Moreover, traditional DEA models rely on linear programming techniques to estimate efficiency scores, which may not adequately capture nonlinear relationships or interactions among inputs and outputs. As a result, the model may overlook certain patterns in the data, leading to biased efficiency estimates. Another significant limitation of traditional DEA is its deterministic nature. Traditional DEA models produce a single efficiency score for each DMU based on the observed input-output data, without accounting for uncertainties or variability inherent in real-world systems. This deterministic approach fails to acknowledge the stochastic nature of many decision-making processes.

With the advent of machine learning techniques, there exists a compelling opportunity to enhance the capabilities of DEA by exploiting the computational power and flexibility offered by these methods. By integrating machine learning algorithms with DEA, researchers could potentially improve the accuracy, robustness, and interpretability of efficiency assessments, thereby advancing the state-of-the-art in performance analysis. In this context, it becomes a scientific duty to create the necessary bridges between machine learning and other fields, such as Data Envelopment Analysis. Machine learning algorithms could complement DEA by providing advanced techniques for, for example, data preprocessing (Chen et al., 2014), variable importance measurement (Valero-Carreras et al., 2024), and the treatment of the curse of dimensionality (Esteve et al., 2023), thereby facilitating more accurate and comprehensive efficiency assessments. Moreover, machine learning models can capture nonlinear relationships and interactions among inputs and outputs, addressing one of the key limitations of traditional DEA approaches.

In the literature, several bridges between machine learning (ML) and Data Envelopment Analysis (DEA) have already been established. However, we have identified certain gaps that we believe our approach introduced in this paper can address. Before mentioning these gaps, we briefly review the main contributions related to ML and DEA. As we are aware, in the literature, there are two predominant streams of research that explore the integration of machine learning with Data Envelopment Analysis[1]. The first stream focuses on adapting existing ML techniques to ensure that the predictive function, typically representing a production function in our context, complies with various shape constraints such as monotonicity or concavity. Researchers in this stream use techniques from ML, such as support vector machines (SVM), neural networks (NN), or decision trees, to develop models that capture the underlying relationships between inputs and outputs by imposing shape constraints on the predictive function. Some of these contributions are the following: Kuosmanen and Johnson (2010) demonstrated the connection between DEA and least-squares regression, introducing Stochastic Non-smooth Envelopment of Data (StoNED). Parmeter and Racine (2013) proposed innovative smooth constrained nonparametric frontier estimators, incorporating production theory axioms. Daouia et al. (2016) introduced a method using constrained polynomial spline smoothing for data envelopment fitting, enhancing precision and smoothness. Esteve et al. (2020) and Aparicio et al. (2021) developed Efficiency Analysis Trees (EAT), improving production frontier estimation through decision trees. Valero-Carreras et al. (2021) introduced Support Vector Frontiers (SVF), adapting Support Vector Regression for

[1]A third line of research in the literature employs Data Envelopment Analysis (DEA) as an alternative method to conventional Machine Learning (ML) classification techniques such as Support Vector Machines (SVM), decision trees, and neural networks. In that line, DEA is utilized to classify observations based on their features instead of measuring technical efficiency. For example, it is applied to identify individuals as carriers of a rare genetic disorder from age and several blood measurements. A recent example of this type of contributions is Jin et al. (2024).

production function estimation. Olesen and Ruggiero (2022) proposed hinging hyperplanes as a nonparametric estimator for production functions. Guerrero et al. (2022) introduced Data Envelopment Analysis-based Machines (DEAM) for estimating polyhedral technologies. Valero-Carreras et al. (2022) adapted SVF for multi-output scenarios, improving efficiency measurement. Guillen et al. (2023a, 2023b, 2023c, 2024) introduced boosting techniques for efficiency estimation in different scenarios. Tsionas et al. (2023) proposed a Bayesian Artificial Neural Network approach for frontier efficiency analysis under shape constraints. Liao et al. (2024) proposed Convex Support Vector Regression (CSVR) to improve predictive accuracy and robustness in nonparametric regression. The second stream of literature adopts a two-stage approach to integrate DEA with ML techniques. In the first stage, researchers apply a pre-existing DEA model, such as the output-oriented radial model, to compute efficiency scores for each observation in the sample (DMUs). In the second stage, the efficiency scores obtained from DEA are treated as the response variable in a 'regression' model based on standard ML techniques (without shape constraints). The original inputs and outputs, along with potentially additional environmental variables, serve as predictor variables in the regression model. By incorporating ML techniques to the performance evaluation framework, researchers aim to develop more robust and accurate predictive models for assessing efficiency. Some of these contributions are the following: Emrouznejad and Shale (2009) explored a novel approach by combining a neural network with Data Envelopment Analysis (DEA) to address the computational challenges posed by large datasets. Liu et al. (2013) compared standard DEA, three-stage DEA, and neural network approaches to measure the technical efficiency of 29 semi-conductor firms in Taiwan. Fallahpour et al. (2016) presented an integrated model for green supplier selection under a fuzzy environment, combining DEA with genetic programming to address the shortcomings of previous DEA models in supplier evaluation. Kwon et al. (2016) explored a novel method of performance measurement and prediction by integrating DEA and neural networks. The study used longitudinal data from Japanese electronics manufacturing firms to show the effectiveness of this combined approach. Aydin and Yurdakul (2020) introduced a three-staged framework utilizing Weighted Stochastic Imprecise Data Envelopment Analysis and ML algorithms to assess the performance of 142 countries against the COVID-19 pandemic. Tayal et al. (2020) presented an integrated framework for identifying sustainable manufacturing layouts using Big Data Analytics, Machine Learning, Hybrid Meta-heuristic and DEA. The paper by Nandy and Singh (2020) presented a hybrid approach utilizing DEA and Machine Learning, specifically the Random Forest (RF) algorithm, to evaluate and predict farm efficiency among paddy producers in rural eastern India. Zhu et al. (2021) proposed a novel approach that combines DEA with ML algorithms to measure and predict the efficiency of Chinese manufacturing companies. Jomthanachai et al. (2021) proposed an integrated method combining Data Envelopment Analysis and Machine Learning for risk management. Boubaker et al. (2023) proposed a novel method for estimating a common set of weights based on regression analysis (such as Tobit, LASSO, and Random Forest regression) for DEA to predict the performance of over 5400 Vietnamese micro, small and medium enterprises. Amirteimoori et al. (2023) introduced a novel modified Fuzzy Undesirable Non-discretionary DEA model combined with artificial intelligence algorithms to analyze environmental efficiency and predict optimal values for inefficient DMUs, focusing on CO_2 emissions in forest management systems. Lin and Lu (2024) presented a novel analytical framework utilizing inverse Data Envelopment Analysis and ML algorithms to evaluate and predict suppliers' performance in a sustainable supply chain context. Omrani et al. (2024) valuated the efficiency of electricity distribution companies (EDCs) from 2011 to 2020 using a combination of DEA, corrected ordinary least squares (COLS), and machine learning techniques. In particular, a three-stage process involving DEA, COLS, support vector regression (SVR), fuzzy triangular numbers, and fuzzy TOPSIS methods are employed, revealing trends in EDC performance and identifying areas needing improvement.

Both streams of research have contributed valuable insights and methodologies for integrating ML with DEA. However, despite these advancements, there remains room for further improvement and innovation in this domain. In this regard, our approach aims to contribute by introducing novel functionalities and complementary methodologies to traditional DEA-based analysis. This is achieved through the exploitation of the favorable properties exhibited by classification models in machine learning. Next, we outline the key contributions of our approach, highlighting its methodological innovations, interpretative advantages, and practical implications for efficiency assessment:

- Methodological Innovation: As we are aware, for the first time in the literature, we propose a classification-based machine learning approach in the second stage of a DEA-ML hybrid framework, moving beyond the conventional regression-based techniques. In the first stage, we employ a standard DEA model to conduct a Pareto-dominance efficiency evaluation, which generates a binary labeling system that distinguishes between efficient and inefficient DMUs. In the second stage, we apply classification models to predict this efficiency label using all available input and output variables.
- Inferential Power: One of the key advantages of our classification-based approach is that it enables the estimation of the probability of a DMU being classified as efficient or inefficient. Unlike traditional DEA models, which primarily serve as descriptive tools, our framework incorporates a probabilistic perspective, allowing researchers and practitioners to infer efficiency status based on statistical learning principles. This

conceptual shift aligns DEA with modern inferential analytics, bridging the gap between efficiency measurement and robust decision-making.

- Reinterpreting DEA as a Classification Problem: Our approach redefines DEA as a classification problem, where the DEA frontier can be interpreted as the separating surface in the variable space between two classes: technically feasible (producible) and infeasible (unproducible) input-output profiles. Under this reinterpretation, efficiency measures can be seen as quantifying the minimum required input and/or output modifications necessary for an inefficient unit to transition from the feasible class to the infeasible class.
- Algorithm-Agnostic Approach for Robust Efficiency Assessments: A key advantage of our framework is its flexibility in algorithm selection. Unlike conventional DEA-ML models that rely on a specific regression technique, our method is not tied to a particular classification algorithm. This flexibility allows us to experiment with various machine learning models—including decision trees, Support Vector Machines, Neural Networks, and ensemble methods—ensuring that the results remain robust and consistent across multiple techniques. Nevertheless, seeking simplicity, in this paper we focus on Neural Networks.
- XAI and Counterfactuals: Another major contribution of our study is the integration of Explainable Artificial Intelligence (XAI), specifically counterfactual methods, into efficiency analysis. Instead of relying solely on conventional DEA scores, we define technical inefficiency for an inefficient DMU as the minimum changes required in inputs and outputs to transition from an inefficient state to an efficient one. These counterfactual-based adjustments offer an intuitive and interpretable way to assess inefficiency.
- Benchmarking with Variable Importance and Directed Projections: Our methodology also introduces a novel benchmarking approach by using the importance ranking of inputs and outputs identified through machine learning models. As highlighted in the literature (e.g., Banker and Morey, 1986; Thanassoulis et al., 2015), understanding the relative importance of variables in efficiency assessments is crucial for strategic decision-making. We propose using this information to assign data-driven weights to inputs and outputs, guiding the projection of inefficient DMUs towards more meaningful and customized improvement paths. This is a significant departure from traditional input- or output-oriented DEA projections, which often rely on arbitrary directional vectors.
- Target Setting Through Counterfactual Benchmarking: The benchmarking framework we propose is further enhanced by incorporating probabilistic efficiency thresholds and applying counterfactual analysis to determine the minimum necessary changes in inputs and outputs that would allow a DMU to be reclassified as efficient. This technique not

only generates concrete improvement targets but also allows practitioners to prioritize adjustments based on their impact on efficiency classification.

- Ranking DMUs and Confidence Thresholds: Expanding on previous works (e.g., Sexton, 1986; Thanassoulis et al., 2008), we propose a novel ranking system for DMUs based on their probabilistic efficiency scores. By exploiting the information provided by directional projections onto the separating surface of the two classes (efficient vs inefficient), we can rank units according to their likelihood of being classified as efficient, providing an interesting evaluation framework compared to traditional DEA-based ranking methods. Additionally, we introduce the concept of confidence-threshold-based peer selection, allowing organizations to tailor reference units based on desired levels of benchmarking certainty.
- Proximity-Based Benchmark Identification: Finally, our approach facilitates a more refined peer selection process by identifying, for each DMU and at every efficiency probability threshold, the closest efficient benchmark unit. This selection is performed using proximity-based metrics, such as Euclidean distance, ensuring that benchmark comparisons are contextually relevant and practically achievable. This enhancement strengthens DEA's applicability by allowing for dynamic and adaptive benchmarking strategies.

The paper is structured as follows: In Section 2, we provide background information on Data Envelopment Analysis (DEA) and the machine learning techniques we will utilize, (Artificial) Neural Networks (NN). Section 3 introduces our novel approach, which integrates DEA with the classification technique, aiming to enhance efficiency assessment for DMUs. We demonstrate the practical implications of this integration and its implications for decision-making through an empirical example based on SABI (Iberian Balance Sheet Analysis System) in Section 4. Section 5 concludes and points out further research lines.

2. Background

This background section provides a concise overview of DEA and the ML technique that we will apply in this paper (in particular, Neural Networks).

2.1. Data Envelopment Analysis

Data Envelopment Analysis (DEA) is a non-parametric method widely used for evaluating the relative efficiency of decision-making units (DMUs) in various fields, including economics, finance, management science and operations research. Introduced by Charnes et al. (1978), DEA offers a powerful framework for assessing the efficiency of DMUs transforming multiple inputs into multiple outputs. DEA operates under the assumption of constant returns to scale (CRS) or variable returns to scale (VRS). VRS is particularly suitable for analysing real-world production processes, where economies of scale may vary across different units.

In this study we evaluate the performance of n observations by measuring their technical efficiency. These observations or DMUs, which could be firms or organizations, utilize m inputs $\mathbf{x}_j = \left(x_{1j},\ldots,x_{mj}\right) \in R_+^m$, such as resources, to generate s outputs $\mathbf{y}_j = \left(y_{1j},\ldots,y_{sj}\right) \in R_+^s$, like goods or services. In this notation, input and output vectors for a specific observation j are presented in bold typeface. In a conceptual framework, the term 'technology' (also called production possibility set) encompasses all feasible input-output combinations. This concept is typically represented as:

$$T = \left\{ \left(\mathbf{x},\mathbf{y}\right) \in R_+^{m+s} : \mathbf{x} \text{ can produce } \mathbf{y} \right\}. \tag{1}$$

Among the non-parametric methodologies utilized to empirically approximate the set T, DEA stands out as one of the most commonly employed approaches in practical applications. Under VRS, Banker et al. (1984) show that that the DEA technology T corresponds to:

$$T_{DEA} = \left\{ \left(\mathbf{x},\mathbf{y}\right) \in R_+^{m+s} : y_r \leq \sum_{j=1}^{n} \lambda_j y_{rj}, \forall r, x_i \geq \sum_{j=1}^{n} \lambda_j x_{ij}, \forall i, \sum_{j=1}^{n} \lambda_j = 1, \lambda_j \geq 0, \forall j \right\}. \tag{2}$$

Numerous technical efficiency measures are available to calculate the technical efficiency of observations with respect to T_{DEA}—for a general definition of these measures see Pastor et al. (2012). In particular, our focus is directed towards a prevalent measure, namely, the Directional Distance Function (DDF) by Chambers et al. (1998). Considering the specific DMU $\left(\boldsymbol{x}_o, \boldsymbol{y}_o\right)$, and a directional vector $\left(\boldsymbol{g}_o^x, \boldsymbol{g}_o^y\right)$ that specifies a direction for projection onto the efficient frontier of T_{DEA}, its technical efficiency can be calculated through the following program.

$$
\vec{D}\left(\boldsymbol{x}_o, \boldsymbol{y}_o; \boldsymbol{g}_o^x, \boldsymbol{g}_o^y\right) = \begin{array}{lll}
\max & \beta & (3.0) \\
s.t. & \sum_{j=1}^{n} \lambda_{jo} x_{ij} \le x_{io} - \beta g_{i0}^x, \quad i = 1,\ldots,m & (3.1) \\
& \sum_{j=1}^{n} \lambda_{jo} y_{rj} \ge y_{ro} + \beta g_{r0}^y, \quad r = 1,\ldots,s & (3.2) \\
& \sum_{j=1}^{n} \lambda_{jo} = 1, & (3.3) \\
& \lambda_{jo} \ge 0, \quad j = 1,\ldots,n & (3.4)
\end{array} \tag{3}
$$

Under this model, a DMU with a score of zero, i.e., $\vec{D}\left(\boldsymbol{x}_o, \boldsymbol{y}_o; \boldsymbol{g}_o^x, \boldsymbol{g}_o^y\right) = \beta^* = 0$ (with * indicating optimality), is considered efficient, signalling that it operates on the efficient frontier of T_{DEA}. Conversely, a value strictly greater than zero, i.e., $\vec{D}\left(\boldsymbol{x}_o, \boldsymbol{y}_o; \boldsymbol{g}_o^x, \boldsymbol{g}_o^y\right) = \beta^* > 0$, implies inefficiency relative to the reference technology T_{DEA}, with a bigger value indicating a worse degree of technical efficiency.

2.2. Neural Networks

In this subsection, we briefly outline the fundamentals of the machine learning technique that will be employed throughout the paper: Neural Networks (NN). NN are a class of learning algorithms inspired by the structure and function of the human brain. They consist of interconnected layers of neurons that process input data through nonlinear transformations to learn complex patterns and relationships. By understanding the underlying principles NN, which determine the label and the probability of belonging to that label, we can harness their capabilities to enhance the DEA methodology.

Neural Networks represent a cornerstone in the field of machine learning, heralded for their ability to learn complex patterns and relationships from data (LeCun et al., 2015; Goodfellow et al., 2016). In this subsection, we briefly delve into the application of Neural Networks in the context of classification tasks, highlighting their versatility, theoretical foundations, and practical implications. Neural Networks are inspired by the structure and function of the human brain, comprising interconnected layers of artificial neurons or nodes. The core principle underlying Neural Networks is the process of propagation, where input data is sequentially passed through multiple layers of neurons, each layer applying a set of weights and activation functions to produce an output. Through an iterative process known as backpropagation, Neural Networks adjust the weights of connections between neurons based on the error between predicted and actual outputs, thereby minimizing a certain loss function and improving predictive accuracy. In this sense, activation functions play a crucial role in Neural Networks by introducing non-linearity into the model, enabling it to capture complex relationships within the data. Common activation functions include the sigmoidal function, hyperbolic tangent (tanh) function, and rectified linear unit (ReLU) function. Each activation function introduces different properties to the model, influencing its ability to learn and generalize from data.

In a binary classification problem, the neural network is designed to distinguish between two possible classes: a positive class ($y=+1$) and a negative class ($y=-1$). The network processes an input x through multiple layers of neurons, applying weighted transformations and nonlinear activation functions. The hidden layers compute intermediate representations:

$$\boldsymbol{h}^{(l)} = f\left(\boldsymbol{W}^{l}\boldsymbol{h}^{(l-1)} + \boldsymbol{b}^{(l)}\right), \tag{4}$$

where $\boldsymbol{W}^{l}$ and $\boldsymbol{b}^{(l)}$ are the weight matrix and bias vector at layer l, and $f(\cdot)$ is a nonlinear activation function. At the final layer, a single neuron outputs a logit z, which is mapped to a probability using the sigmoid function:

$$P(y=+1 \mid \boldsymbol{x}) = \sigma(z) = \frac{1}{1+e^{-z}}. \tag{5}$$

This probability represents the likelihood that the given input belongs to the positive class ($y=+1$). Since there are only two possible outcomes, the probability of the negative class ($y=-1$) is simply: $P(y=-1 \mid \boldsymbol{x}) = 1 - P(y=+1 \mid \boldsymbol{x})$.

The performance of Neural Networks hinges on the selection of hyperparameters such as the number of layers, the number of neurons per layer, learning rate, and regularization parameters. Hyperparameter tuning is essential to optimize model performance and prevent issues like overfitting or underfitting.

An illustrative example of the configuration of a neural network in the context of a binary classification problem, with two predictor variables, would consist of two neurons in the input layer, reflecting the number of variables involved in the model. In the output layer, a single neuron would be located to assign the corresponding class to each observation. Between these layers lies, in this case, one hidden layer, composed of three neurons. Figure 1 depicts the structure of this neural network with a configuration of 2-3-1.

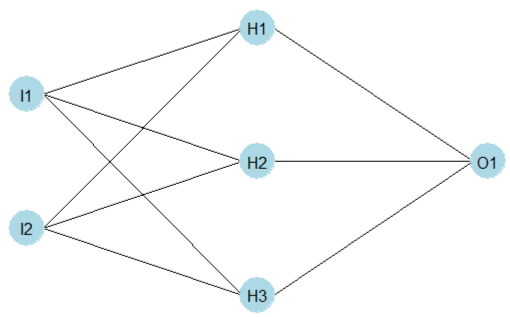

Figure 1. An example of a very simple artificial Neural Network

Figure 2 illustrates the nonlinear decision boundary (separating surface) generated by the neural network, which partitions the feature space into two distinct regions. One region is associated with class $y=1$, while the other corresponds to class $y=0$. This separating surface emerges as a result of the network's learned transformations, effectively capturing complex patterns in the data that a linear classifier would fail to model. The figure provides a visual representation of how the network adapts to the underlying structure of the dataset, demonstrating its ability to perform nonlinear classification.

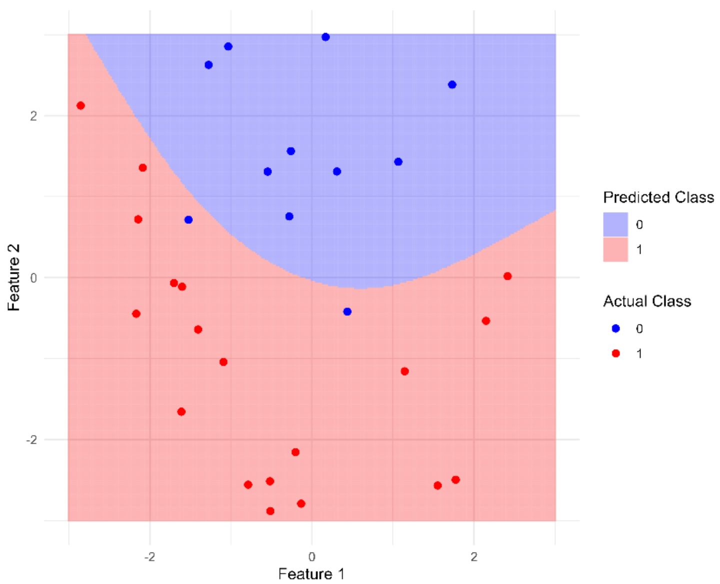

Figure 2. Nonlinear decision boundary generated by a Neural Network.

2.3. eXplainable Artificial Intelligence

The so-called eXplainable Artificial Intelligence (XAI) has emerged as a critical area of research aimed at enhancing the transparency, interpretability, and trustworthiness of machine learning models (Wachter et al., 2017). In this section, we provide an overview of XAI principles and delve into the concept of counterfactual methods, a subset of XAI techniques that facilitate insightful explanations of model predictions.

Overall, XAI encompasses a diverse set of methodologies and techniques designed to elucidate the decision-making process of machine learning models. As AI (Artificial Intelligence) systems become increasingly complex and ubiquitous, there is a growing need for transparency and interpretability to foster trust and facilitate human understanding of model behaviour. XAI aims to address this need by providing explanations that are understandable, intuitive, and actionable for end-users, stakeholders, and domain experts. To address this, XAI provides methodologies designed to clarify the decision-making process of ML models. These approaches aim to generate explanations that are understandable, actionable, and intuitive for end-users and stakeholders, enabling better model validation and facilitating the identification of relationships within the data.

2.3.1 Counterfactual Explanations

In particular, counterfactual methods represent a prominent approach within the realm of XAI, focusing on the generation of alternative scenarios or 'counterfactuals' to explain model predictions. The fundamental concept underlying counterfactual methods is the creation of hypothetical instances that are similar to the observed data but differ in one or more attributes. By systematically altering the features of a given instance and observing the corresponding changes in model predictions, counterfactual methods provide valuable insights into the factors driving model decisions and predictions. Moreover, counterfactual explanations offer intuitive and interpretable insights into machine learning models by highlighting the causal relationships between features and model outcomes. These explanations typically take the form of 'what-if' scenarios, where adjustments are made to features to generate counterfactual instances that lead to desired outcomes. By identifying the minimal changes required to alter a model prediction, counterfactual explanations shed light on the underlying decision-making process and enable decision-makers to understand the model's behaviour in specific contexts (for example, in our production context the question could be 'What is the minimum amount of adjustment in inputs and/or outputs that a technically inefficient DMU would need to undertake to transition into being considered efficient with a certain probability?'). Thus, the counterfactual method involves projecting an observation from one class onto the separating surface of the two classes, meaning the projection stops just before a change in label occurs. This 'projection' strategy will be incorporated into our approach in this paper to measure technical efficiency in the context of machine learning and efficiency analysis (see Section 3).

2.3.2 Feature Significance Analysis and Sensitivity Analysis

To complement counterfactual analysis, we incorporate feature significance analysis, focusing on understanding the contribution variables to the model's predictions. There are several approaches to feature significance analysis, such as rule extraction methods (see, for example, Tickle et al., 1998; Fogel & Robinson, 2003; Martens et al., 2007), visualization techniques (see, for example, Craven & Shavlik, 1992; Tzeng & Ma, 2005; Cho et al., 2008), Sensitivity Analysis (SA) (Ruck et al., 1990), and more recent methods such as SHAP (Lundberg & Lee, 2017) and LIME (Ribeiro et al., 2016). We decided to use SA for several reasons. First, extraction rules typically simplify the model's complexity to produce more understandable rules, which involves discretizing the classifier, leading to information loss, and failing to accurately represent the original model. Instead, SA is a straightforward method that treats the original fitted model, querying it with sensitivity samples and recording the corresponding outputs. Second, visualization techniques are often designed for a specific machine learning method limiting their general applicability. This is a disadvantage compared to SA, which can be applied to any supervised machine learning method. Third, methods like SHAP and LIME are computationally intensive and more challenging to interpret, especially in high-dimensional datasets. In contrast, SA is computationally efficient, simple to implement, and provides clear, actionable insights, making it a practical choice for decision-makers. In particular, SA works by perturbing each variable across its range while keeping other variables constant at their baseline values.

3. Integrating ML techniques for classification and Data Envelopment Analysis

In this section, we perform the integration of machine learning techniques for classification tasks with Data Envelopment Analysis (DEA) to enhance the measurement of technical efficiency. By combining the strengths of both methodologies, we aim to provide robust and insightful efficiency assessments of a set of DMUs. In this case, while other ML classification methods could be considered, we focus here on Neural Networks.

3.1. Reinterpreting DEA as a Classification Machine Learning Technique and its way to measure technical efficiency as a XAI method

Before introducing our methodology, we aim to elucidate the reinterpretation of DEA, through a graphical toy example (Figure 3), as a classification method that also resorts to counterfactual analysis to measure technical inefficiency. DEA can be conceptualized as a classification model wherein the two classes (labels) represent feasible and infeasible input-output profiles, with the production frontier delineating the separating surface and efficient units positioned precisely onto this surface. Furthermore, within the feasible but inefficient set of DMUs, this reinterpretation implies that the typical efficiency measures utilized in DEA can be reinterpreted within the realm of eXplainable Artificial Intelligence (XAI) principles, particularly in relation to the notion of counterfactual scenarios. Specifically, the movement of an inefficient DMU, by improving its observed inputs and/or outputs in accordance with the orientation and type of efficiency measure selected (e.g., using the Directional Distance Function, model (3)), signifies transition within its original class 'feasible' toward the exact threshold where any further minimal change would result in the unit being classified as 'unfeasible' (through its projection onto the efficient frontier, i.e., the separating surface). This movement resembles a counterfactual that quantifies the level of technical inefficiency within the ‘feasible’ class through DEA, thus highlighting the conceptual linkage between DEA and XAI principles.

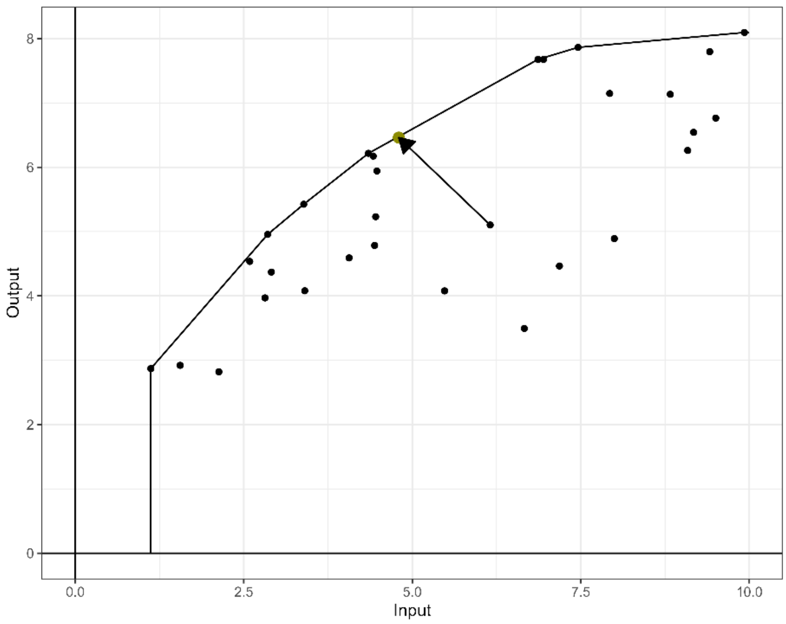

Figure 3. Data Envelopment Analysis and the Directional Distance Function

After drawing a parallel between standard DEA approaches and classification ML methods, showing that DEA efficiency measures can be considered as a specific case of XAI, we now proceed to introduce our method.

3.2. Classifying DMUs by their (in)efficiency class and measuring technical efficiency

The core concept underlying our model is a multi-stage methodology aimed at enhancing efficiency assessment through the fusion of DEA and ML techniques. Our approach operates in three distinct phases: Firstly, we employ standard DEA to categorize DMUs into efficient and inefficient categories. Subsequently, in the second phase, we address the challenge of class imbalance (efficient vs inefficient). In the third phase, we employ a classification ML model, wherein the response variable is the efficiency status, and the classification features include both inputs and outputs. Finally, in the fourth phase of our approach, we ascertain a robust measure of technical inefficiency through the application of XAI principles. Specifically, given a model measuring technical efficiency (such as the output-oriented radial model), we determine the minimum increase required in the output of each inefficient DMU to transition its class from inefficient to efficient.

Next, we introduce our approach in the form of an algorithm with different steps:

Step 1 [Labeled data process]: Utilize the additive DEA model (Charnes et al., 1985), model (4), to partition the set of DMUs into two categories (efficient vs inefficient) based on the optimal value of the optimization program. A value of zero indicates that the evaluated unit is not Pareto-dominated by any technically feasible input-output combination within the standard DEA production possibility set. This condition underscores the efficiency of the evaluated unit, demonstrating that there is no room in the observed sample for enhancing any input and/or output without compromising the feasibility of the unit under assessment.

$$
\left.\begin{array}{llll}
A_{DEA}\left(\boldsymbol{x}_o, \boldsymbol{y}_o\right) = & \max & \sum_{i=1}^{m} s_{io}^{-} + \sum_{r=1}^{s} s_{ro}^{+} & (6.0) \\
& s.t. & \sum_{j=1}^{n} \lambda_{jo} x_{ij} = x_{io} - s_{io}^{-}, \quad i = 1,...,m & (6.1) \\
& & \sum_{j=1}^{n} \lambda_{jo} y_{rj} = y_{ro} + s_{ro}^{+}, \quad r = 1,...,s & (6.2) \\
& & \sum_{j=1}^{n} \lambda_{jo} = 1, & (6.3) \\
& & \lambda_{jo} \geq 0, \quad j = 1,...,n & (6.4) \\
& & s_{io}^{-}, s_{ro}^{+} \geq 0, \quad \forall i, \forall r & (6.5)
\end{array}\right\} \quad (6)
$$

If, $A_{DEA}(x_o, y_o) > 0$, then DMU (x_o, y_o) is (technically) inefficient. The set of all inefficient DMUs is denoted as I. Otherwise, that is, if $A_{DEA}(x_o, y_o) = 0$, then DMU (x_o, y_o) is (technically) efficient. The set of all efficient DMUs is denotes as E.

Step 2 [Class balancing phase]: Addressing the challenge of class imbalance (efficient and inefficient) is crucial for prediction by means of ML techniques (see, for example, He & Garcia, 2009). Imbalanced datasets often compromise the performance of standard algorithms, favouring the majority class and neglecting the minority class. In our production context, datasets typically exhibit a higher proportion of inefficient units, which can skew model outcomes and adversely affect the accuracy of predictions. To address this issue, we adopt a modified version of the Synthetic Minority Oversampling Technique (SMOTE) (Chawla et al., 2002) to generate synthetic examples of the minority class (efficient units). This adaptation allows us to tailor the synthetic data generation process to better fit the characteristics of our dataset and context. Next, we describe the specific implementation process of our adapted approach to generate synthetic units.

First, we determined the necessary number of synthetic units to balance the proportion of units in both classes (efficient vs. inefficient units). There is no exact proportion that guarantees an ideal balance in the dataset. Weiss and Provost (2003) suggest testing performance with different percentages of minority class examples to identify the optimal class distribution, or an optimal range of class distribution, for the training set. They conclude that the proportion of the minority class should ideally fall between 20% and 40%. Like them, in our approach, we test the performance of 20%, 25%, 30%, 35% and 40% and generate the synthetic units for each scenario.

Step 2a: Let us suppose that we select a particular balance level. Then, we achieve balance by increasing the number of efficient DMUs, which is the most common approach in our production context. Specifically, we generate convex combinations of $m + s$ among the DMUs labelled as efficient in Step 1. The total number of combinations is calculated as $\binom{n_E}{m+s}$, where n_E is the number of efficient DMUs and k is the sum of the number of inputs and outputs, $m + s$. For each combination, a synthetic unit is generated by applying the same weights to the DMUs involved in that combination. The weight is defined as $\upsilon = \frac{1}{m+s}$. Once all convex combinations have been created, we use the additive DEA model (4) to identify which of these combinations are Pareto-efficient. If the number of synthetic units remains insufficient, additional random DMUs are generated based on combinations where we found an efficient synthetic point. In this process, the weights are randomly selected within the range [0.05, 0.95] to ensure that no weight is equal to zero. To maintain consistency and ensure that the sum of all weights equals 1, each weight is normalized by dividing it by the total sum of all weights, yielding a new relative weight for each DMU. When the pre-fixed balance level is achieved, the generation of synthetic units stops. If none of the $\binom{n_E}{m+s}$ combinations yield a Pareto-efficient point, we proceed by selecting combinations of $m+s-1$ efficient DMUs. If this approach also fails, we reduce the number to $m+s-2$ and continue iterating in this manner until a solution is found.

Step 2b: Additional synthetic inefficient DMUs are generated following a procedure similar to the one described in the previous paragraph. The process consists of four key steps. First, synthetic convex combinations are created using equal weights. Second, the additive DEA model (Equation (6)) is applied to determine which of these convex combinations are inefficient. Third, a large random sample of inefficient convex combinations—for example, 20 times the desired number of synthetic inefficient units—is selected. Fourth, the sampled units are grouped into quantiles based on their slack values, and a balanced subset is obtained by randomly selecting an equal number of units from each quantile until the target distribution is achieved. This systematic approach ensures a well-distributed set of synthetic inefficient DMUs while maintaining representativeness across different inefficiency levels.

Step 3 [Fitting the ML model]: In this phase, a classification ML model is implemented where the dependent variable denotes the efficiency status (efficient [class +1] vs. inefficient [class -1]), while the independent variables (features) comprise all inputs and outputs. In this step, the parameters of the ML model will also be fine-tuned through cross-validation, ensuring the determination of an optimal parameter configuration, an ideal balance rate and a final classification model $\Gamma(x, y)$. The best balance rate is selected by comparing the model's performance among a grid of possibilities: 20%-40%. If the original dataset is large, we propose creating training, testing, and validation partitions to evaluate how the model interacts with data not used during the fitting phase. If dividing our data into three partitions is not the best option, we propose testing the model's performance on the original dataset and selecting the balance level that provides the best results. In case multiple models exhibit equal performance, the balance level will be determined based on the model's performance with its respective balanced dataset. If equality persists, the smallest balance level will be selected, following the principle of parsimony. Finally, the best $\Gamma(x,y)$ predicts the probability of belonging for each class and classifies the input-output bundle (x,y) as (technically) efficient (+1) or inefficient (-1) through the rule: if $\Gamma(x,y)>0.5$, then (x,y) is classified as efficient; otherwise, (x,y) is classified as inefficient.

Step 4 [Measuring technical inefficiency]: Select a standard technical efficiency measure. For example, the DDF simultaneously captures improvements across input-output dimensions. The objective of this measure is to reduce inputs, while increasing outputs. The directional vector is defined as $(\boldsymbol{g}^x, \boldsymbol{g}^y)$, where $\boldsymbol{g}^x$ represents the relative importance of each input multiplied by its respective observed mean, and $\boldsymbol{g}^y$ represents the relative importance of each output, multiplied by its respective observed mean. The relative importance of each variable is calculated through the SA method (see Section 2.3.2). As we are aware, the selection of this type of directional vector is also original (Wang et al., 2019). This choice of directional vector is particularly noteworthy for several reasons. First, incorporating the relative importance of input and output variables—derived from their contribution to predicting efficiency status—introduces a novel and objective criterion for defining the direction of improvement. Unlike traditional approaches that often rely on arbitrary or uniform weightings, this method exploits data-driven insights obtained through sensitivity analysis. Second, using the same directional vector to evaluate inefficiency across all DMUs ensures that the optimal β value remains comparable across units. This uniformity enhances the interpretability of inefficiency scores, as differences in inefficiency are not influenced by unit-specific directional choices but rather reflect genuine performance gaps relative to a common benchmark. Third, by constructing the directional vector based on the mean values of inputs and outputs, we ensure that the measure is inherently dependent on the original units of measurement. This property makes the inefficiency measure unit-invariant, meaning that results remain consistent regardless of the scale or units in which the inputs and outputs are expressed. This avoids potential distortions that could arise from differences in variable magnitudes or measurement scales.

Next, the unknown value β^* (i.e., the level of technical inefficiency) is determined using contrafactual analysis. In particular, we should answer the question: "What is the minimum modification required for the evaluated DMU to be classified as efficient with at least probability p?". The corresponding projection (input-output targets) is calculated through (7) where $\boldsymbol{x}_i$ and

$\boldsymbol{y}_i$ are the observed inputs and outputs for DMU_i, respectively, and β takes values within a predefined grid:

$$\left(\boldsymbol{x}_i - \beta \boldsymbol{g}^x, \boldsymbol{y}_i + \beta \boldsymbol{g}^y\right). \quad (7)$$

If the target probability p lies between two consecutive β values, we refine the search by iteratively testing new β values within that range and recalculating their associated probabilities $\Gamma\left(\boldsymbol{x}_i - \beta \boldsymbol{g}^x, \boldsymbol{y}_i + \beta \boldsymbol{g}^y\right)$. When the algorithm converges to the probability that meets the desired confidence threshold, the corresponding β value is recorded as the minimum adjustment required for the DMU to achieve efficiency. If the efficiency probability of a DMU observed exceeds the pre-fixed threshold p, the β value will be set to 0, and the projection will coincide with the observed DMU. Moreover, our algorithm does not consider projections with input values less than the minimum observed. As a result, some DMUs could never reach the established threshold p. In such cases, we use as β^* the biggest feasible β in the considered grid.

Finally, given a predefined threshold p, the peers for each evaluated DMU are identified as the closest unit classified as efficient at level p to the assessed DMU. This selection ensures that the chosen peer represents the most comparable efficient unit, providing a meaningful benchmark for performance evaluation. By resorting to proximity within the input-output space, the approach facilitates a more intuitive interpretation of efficiency adjustments required for the assessed DMU to reach the predefined efficiency level. Note that peers can change for each DMU depending on p. Formally, for DMU $\left(\boldsymbol{x}_i, \boldsymbol{y}_i\right)$, and given probability threshold p, its corresponding peer is determined as $\left(\boldsymbol{x}_{i^*}, \boldsymbol{y}_{i^*}\right) = \arg\min_{k=1,\ldots,n} \left\{ \left\| \left(\boldsymbol{x}_i, \boldsymbol{y}_i\right) - \left(\boldsymbol{x}_k, \boldsymbol{y}_k\right) \right\|_2 : \Gamma\left(\boldsymbol{x}_k, \boldsymbol{y}_k\right) \geq p \right\}$. In case of a tie, we randomly select one of the possible solutions.

3.3. An illustrative example.

Next, we will illustrate our method through a numerical example, complemented by several figures. As the classification ML model, we employ Neural Networks (NN).

In this example, we utilize a simulated dataset made up of 40 DMUs (D) that use a single input to produce a single output. Following the algorithm, step 1 labels the available data according to the additive model through standard DEA. In this example, 3 DMUs are efficient with all their optimal slacks in model (4) equal to 0. The remaining 37 are marked as 'inefficient'. The efficient DMUs are: 6, 7 and 31 (see Figure 4). In this case, there is an imbalance in the labels, with 7.5% of the units being efficient and 92.5% being inefficient.

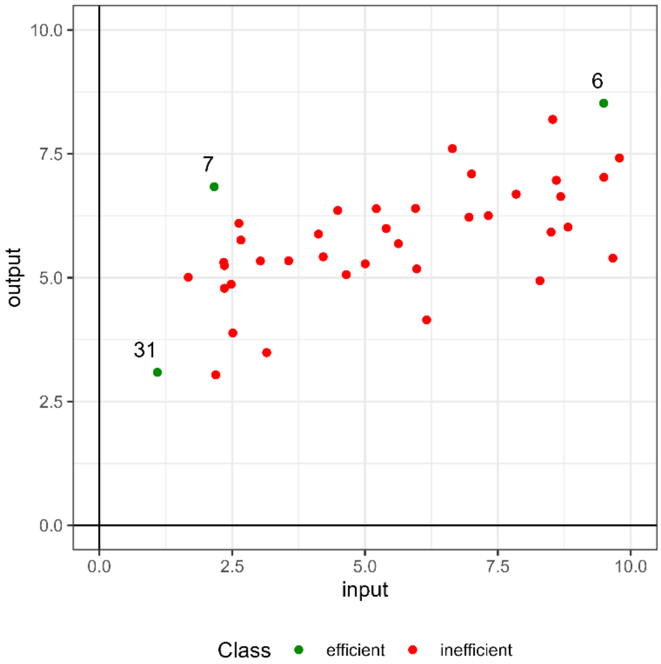

Figure 4. Labeling through the standard DEA additive model.

Step 2 of the method balances the dataset by generating synthetic units. In this example, the minority class consists of efficient DMUs, which are augmented to achieve balance. The procedure for creating new efficient synthetic units (set $\hat{E}$). Figure 5 illustrates the augmented dataset with a 25% minority class level. Initially, there were only 3 observations labelled as 'efficient', which increased to 13 after the creation of synthetic efficient units. Once the data imbalance has been addressed, the dataset consists of $\widehat{D} = D \cup \hat{E}$ with 50 units, with an approximately 1:4 ratio between units labelled as 'efficient' and 'inefficient'. Nevertheless, this serves only as an illustration of the outcomes obtained when applying the balancing process. In practice, a neural network (NN) is required to determine the optimal class balancing level within the 20%-40% range.

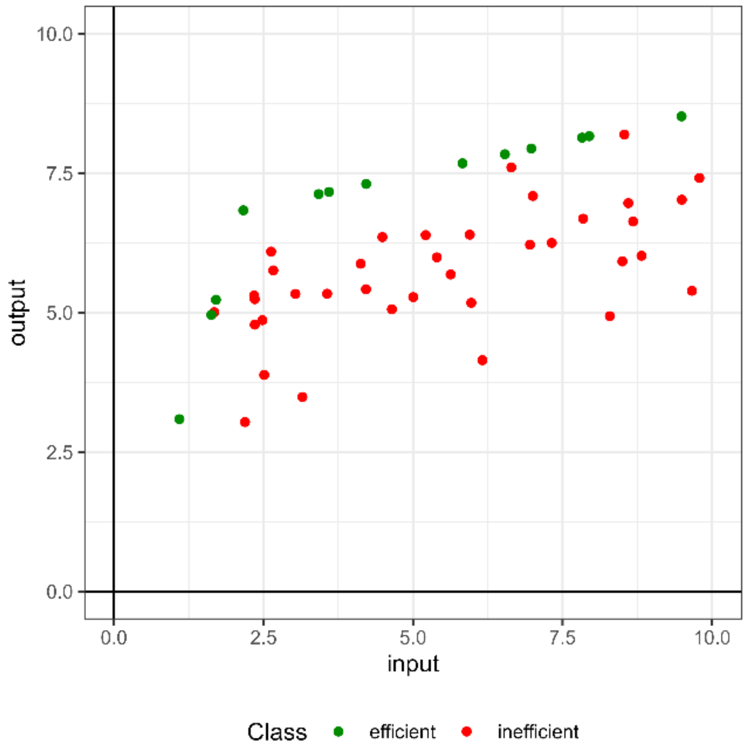

Figure 5. The labeled dataset that will be used for model training.

The third step involves training the NN machine learning model. Seeking simplicity, we employ a neural network (NN) with an input layer containing as many nodes as input variables, a single hidden layer, and a final output layer consisting of a single neuron. In practice, we use the R package caret (Kuhn, 2008) to facilitate model training, specifically using the NN implementation from the nnet package (Venables & Ripley, 2002). A grid is defined with selected hyperparameters for model fitting size (the number or neurons in the hidden layer) (1, 5, 10, 20 and 30) and decay parameter (a regularization constraint) (0, 0.1, 0.01, 0.001 and 0.0001). To determine these hyperparameters, a 5-fold cross-validation was implemented. Another hyperparameter for us is the class balancing level (0.2, 0.25, 0.3, 0.35 and 0.4). In this sense, Table 1 presents the performance of the fitted models for their respective balanced datasets. We use standard metrics commonly applied in ML problems. Our focus is on metrics related to the 'efficient' class, such as sensitivity (the proportion of actual positives correctly identified, or true positive rate), precision (the proportion of positive predictions that are actually correct, or positive predictive value), F1 score (the harmonic mean of precision and sensitivity, balancing detection accuracy and reliability), and balanced accuracy (the average of true predictions for each class). Due to the limited number of DMUs, a validation partition was not created. Performance was evaluated using the observed data, which remains consistent across all models. Since there is a tie for balance levels 0.25, 0.35, and 0.4, we consider the performance using the whole dataset (real and synthetic units), which differs by the balance level established. After this evaluation, 0.25 and 0.4 remained tied. Following the principle of parsimony, we selected the 0.25 imbalance dataset. After adjusting the model, the optimal hyperparameters for this dataset were: size = 20 and decay = 0.

Step 1: Performance using real dataset

Balance	Sensitivity	Precision	F1	Balanced accuracy
0.25	1	1	1	1
0.35	1	1	1	1
0.4	1	1	1	1
0.3	1	0.75	0.86	0.99
0.2	0.67	1	0.8	0.83

Step 2: Performance using train dataset

Balance	Sensitivity	Precision	F1	Balanced accuracy
0.25	1	1	1	1
0.4	1	1	1	1
0.35	1	0.95	0.98	0.99
0.3	1	0.94	0.97	0.99
0.2	0.7	1	0.82	0.85

Table. 1 Performance results for each fitted model across different class balancing levels.

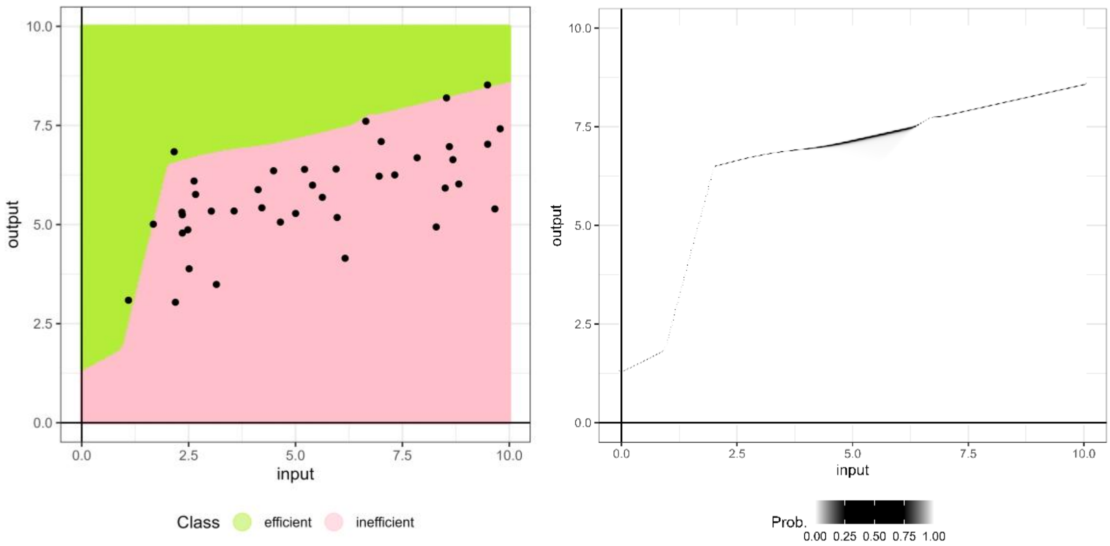

Figure 6. On the left, the predicted regions generated by the new approach are displayed alongside the original unlabeled DMUs. On the right, the uncertainty regions are shaded in black as predicted by the fitted model, with certainty regions shown in white.

Figure 6 (on the left) displays the separating hyperplane generated by the trained model at probability 0.5. Because it is not possible to visualize the hyperplane directly, we create a grid ranging from 0 to 10 in both dimensions and predict the class for each point. Green points represent those with a probability higher than 0.5, classified as 'efficient,' while the red points correspond to those with a probability of 0.5 or lower, classified as 'inefficient. DMUs (black points) are classified using this method. Figure 6 (on the right) represents uncertainty through a gradient based on the probabilities predicted by the model. Areas with a predicted probability between 0.25 and 0.75 are shaded in black, indicating maximum uncertainty, while regions with probabilities closer to 0 or 1 are displayed in white, reflecting greater certainty.

To determine β^* (the DDF), the input and output target and peers for each DMU, we perform a SA analysis (see Section 2.3.2) using the Rminer library (Cortez, 2010) and we find out that the model considers the output variable to be twice as important as the input variable when classifying a DMU as efficient or inefficient. The SA result for the input is $SA_x = 0.333$ and for the output $SA_y =$ 0.667. After SA analysis result, we define the directional vector as $\left(\boldsymbol{g}_x, \boldsymbol{g}_y\right) = \left(SA_x \cdot \bar{x}, SA_y \cdot \bar{y}\right) = (1.804, 3.848)$

Figure 7 illustrates the projection of DMU 22, which is classified as inefficient with an input value of 4.49 and an output value of 6.36 with $\Gamma\left(DMU_{22}\right)=0$. Using the above directoral vector, we calculate the minimum β required to reach the specified efficiency confidence level, such as 0.75. The resulting projection reduces the input to 4.17 (its input target) and increases the output to 7.03 (its output target), corresponding to a $\beta^{*}=0.17$.

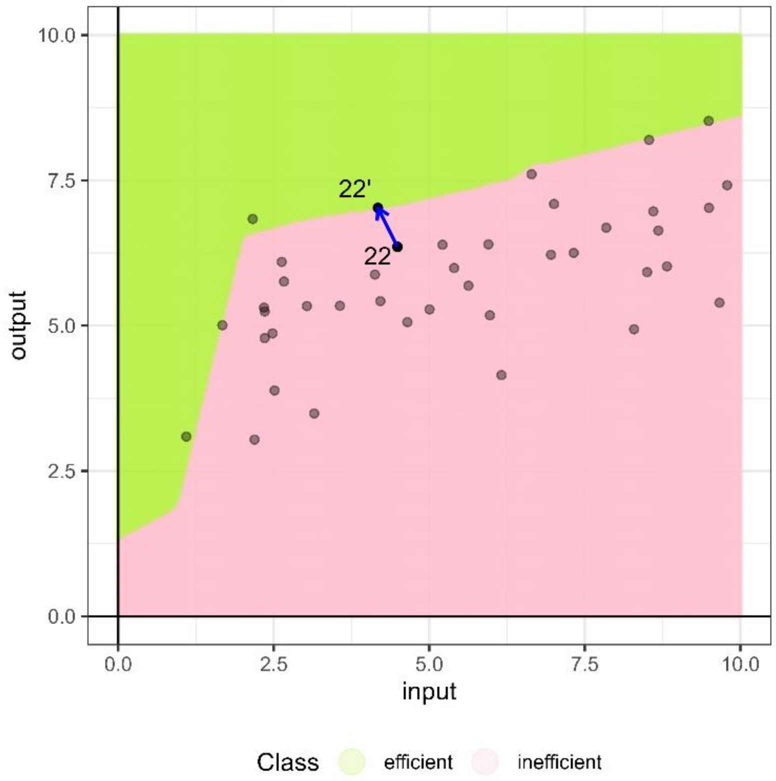

Figure 7. Projection of DMU 22 at an efficiency confidence level of p=0.75.

For this confidence level, only three DMUs have a probability greater than 0.75: 6, 7, and 31. These are the same DMUs labeled in Step 1. For DMU_{22}, the efficient peer of at least 0.75 is DMU_7.

4. An empirical application: Efficiency Assessment of the Valencian Food Industry

In this section, we focus on applying the new approach to real-world data from Spain's food industry, a sector that plays a crucial role in the country's economy. The food industry in Spain is both economically significant and culturally rich, seamlessly combining traditional practices with modern technological advancements. Its scope covers the entire food value chain, transforming raw agricultural products into a wide variety of food items consumed domestically and exported internationally. This industry is supported by a diverse ecosystem, ranging from small-scale farmers dedicated to preserving heritage techniques to large companies adopting advanced production systems. It is also a vital source of employment, spanning agriculture, processing, distribution, and retail. Numerous studies worldwide have analyzed efficiency in food industries, with examples including India (Kumar and Basu, 2008), Mexico (Flegl et al., 2022), Taiwan (Dadura and Lee, 2011), and Indonesia (Machmud et al., 2019). Such analyses provide valuable insights into the operational dynamics of food sectors across different regions.

In Spain, the economic environment is shaped by its 17 autonomous communities, each characterized by distinct policies and market conditions. This regional diversity introduces significant complexity into any economic analysis, as variations in regulations and economic frameworks influence business operations at the community level. The Valencian Community, selected as the focus of this study, exemplifies such diversity. Known for its strong agricultural exports and medium-sized enterprises, this region provides a representative case for evaluating efficiency in the Spanish food industry.

The dataset used for this analysis consists of 97 food industry companies located in the Valencian Community, each employing more than 50 workers, and collected from the SABI[2] database for year 2023. The dataset includes several variables that comprehensively reflect the operational and financial profiles of the companies. The output variable, operating income (in millions of Euros), captures revenue generated from core business activities. Input variables include total assets (in millions of Euros), representing the resources utilized; the number of employees, indicating workforce size; tangible fixed assets (in millions of Euros), such as buildings and machinery essential for production; and personnel expenses (in millions of Euros), encompassing costs like salaries, benefits, and training. Together, these variables enable a detailed examination of resource allocation, labor engagement, and financial investments, forming the basis for a robust analysis of operational efficiency within the Valencian food industry (see Table 2). To better understand the characteristics of the dataset and the challenges it presents for analysis, Table 2 presents the descriptive statistics for the sample. Examining the data, we observe that the dataset includes both very small and very large companies. The maximum and minimum values are significantly distant from the mean and median, highlighting the wide dispersion in the data. This dispersion affects the central tendency measures, resulting in a notable difference between the mean and median.

[2] We resorted to the SABI (Iberian Balance Sheet Analysis System) database: https://sabi.bvdinfo.com.

Building upon this production framework, we will employ the technique described in this paper, which combines ML techniques for classification and DEA, to determine a robust technical efficiency analysis. This approach should allow us to capture the complex intricacies and idiosyncrasies of the food industry in the Valencian Community, providing a more accurate and contextualized perspective on efficiency.

	Inputs				Output
	Total assets	Employees	Fixed assets	Personal expenses	Operating income
Min.	1.537	50	0.142	1.037	2.382
1st Qu.	8.989	75	2.680	2.167	12.994
Median	24.555	98	6.258	3.059	29.138
Mean	41.030	201	15.280	6.757	62.307
3rd Qu.	52.409	240	20.096	8.244	72.688
Max.	258.825	1076	140.689	36.789	460.578

Table. 2 Main statistics metrics.

In this dataset, the additive model identified 15 DMUs out of the 97 observations as efficient. After labelling the data, the next step involves balancing the dataset and fine-tuning the NN. Regarding the balance, since 15.43% of the DMUs are labelled as efficient, which is a value closed to the minimum permissible percentage (20%), we compare the performance of the model without balancing to scenarios where the efficient class is balanced between 20% and 40% at increments of 5%. In each scenario, we fine-tune a NN with a unique hidden layer. A grid of selected hyperparameters is defined for model fitting, including $size(1,5,10,20 \text{and} 30)$ and $decay(0,0.1,0.01,0.001,0.0001)$. As shown in Table 3, the best performance is achieved in the scenario with 40% imbalance, where the optimal tuning consists of a hidden layer with 20 units and a decay of 0.01. Notably, the second-best performance in terms of balanced accuracy is observed when the dataset is not balanced (0% scenario). However, this model has the worst precision while achieving the highest sensitivity, indicating that it tends to overclassify DMUs as efficient, resulting in less reliable predictions.

Balancing the dataset improves precision at the expense of sensitivity, as observed in the scenarios with 20% and 40% balance, where precision reaches 1. Among these, the scenario with 40%

balance demonstrates the best overall performance across all metrics, striking a better equilibrium between precision, sensitivity, F1 score, and balanced accuracy.

Balance	Sensitivity	Precision	F1	Balanced accuracy
0.4	0.93	1	0.97	0.97
0	1	0.60	0.75	0.94
0.2	0.87	1	0.93	0.93
0.3	0.73	0.85	0.79	0.85
0.35	0.73	0.85	0.79	0.85
0.25	0.60	0.82	0.69	0.79

Table. 3 Performance results of different models depending on balancing levels

After selecting the best-performing model, we are able to predict the probability of being efficient for each DMU. A total of 14 DMUs are predicted to have a probability of efficiency exceeding 0.5, one less than initially labelled in the first step. This result highlights the flexibility of the ML model, which, in addressing the complexity of the dataset, classifies one observation as inefficient to prioritize robust and reliable outcomes over strict adherence to initial labels.

Next, we perform SA on the selected model. The relative importance of the variables used by the model to assign probabilities of efficiency is presented in Figure 9. Operating income accounts for 50.8% of the total importance, highlighting its dominant role in the model's decision-making process. Among the input variables, the relative importance is distributed as follows: employees (25.4%), total assets (12.5%), fixed assets (11.2%), and personal expenses (0.1%). These relative importance results are subsequently used to define the directional vector as $(5.129, 51.054, 1.711, 0.007, 31.652)$ and calculated the β^* value for each DMU through counterfactual analysis, as explained in Section 3.

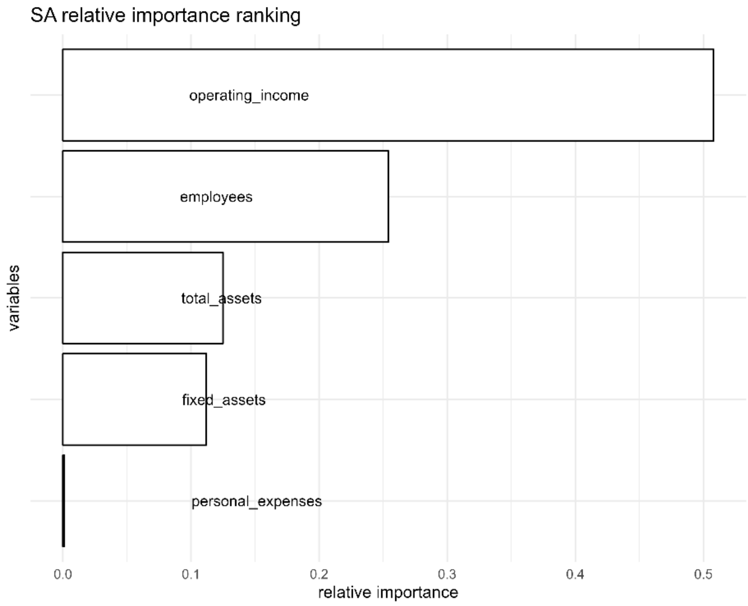

Figure 8. Relative importance of variables, ordered by significance.

Additionally, in Table 4, we rank the DMUs in the sample based on their probabilities of being efficient estimated through the fitted NN. In particular, we show the top 25. Moreover, we identify peers for each DMU using several thresholds higher than or equal to 0.75. The first 14 DMUs exhibit probabilities very close to 1, while DMUs ranked 15 and 16 have probabilities near 0.5, and the remaining DMUs show probabilities close to 0. Notably, DMU 26, originally labelled as efficient, is now classified as inefficient by the model, demonstrating its ability to provide extra insights in comparison with the traditional approach. In the peer columns, we indicate the reference DMU for each unit. In the 0.95 scenario, many DMUs, from DMU 9 onward, change their peer to one among the top 9, all of which have probabilities exceeding 0.95. This is because, beyond DMU 9, no other DMUs have probabilities higher than 0.95. Across the entire dataset, there are the same 14 peers in both the 0.75 and 0.85 scenarios, while only 9 peers remain in the 0.95 scenario.

Ranking	DMU	Probability of being efficient	Peer 0.75	Peer 0.85	Peer 0.95
1	2	0.9999	2	2	2
2	18	0.9998	18	18	18
3	3	0.9996	3	3	3
4	17	0.9983	17	17	17
5	20	0.9962	20	20	20

6	36	0.9960	36	36	36
7	46	0.9894	46	46	46
8	1	0.9868	1	1	1
9	56	0.9705	56	56	56
10	62	0.9486	62	62	46
11	93	0.9441	93	93	46
12	92	0.9335	92	92	46
13	9	0.9288	9	9	3
14	97	0.9176	97	97	46
15	25	0.4981	17	17	17
16	26	0.4909	17	17	17
17	91	0.0735	93	93	56
18	22	0.0560	18	18	18
19	43	0.0549	46	46	46
20	85	0.0490	93	93	46
21	95	0.0338	56	56	56
22	83	0.0335	92	92	46
23	44	0.0183	56	56	56
24	75	0.0167	62	62	46
25	94	0.0099	93	93	46

Table 4. Top 25 DMUs ranked by probability of efficiency and their corresponding peers.

Additionally, we compute the main statistics for the efficient projections in each scenario to observe changes in the metrics. For the thresholds of 0.75, 0.85, and 0.95, 25, 31, and 37 DMUs, respectively, fail to achieve the established probability level due to the constraints discussed in Section 3 (in particular, reducing any input below the minimum observed value in each dimension is not permitted). To further analyze these results, Tables 5, 6, and 7 summarize the mean, median, and standard deviation for these scenarios, highlighting the impact of increasing probability thresholds on the projections (in particular, on the input and output targets). The percentage increments are shown in brackets for better interpretation. According to Table 5, if all DMUs were to adjust following our director vector to be considered efficient, total assets would, on average, decrease by 13% to 15%, employees by 26% to 30%, and fixed assets by 12% to 13%. Operating income would increase by 53% to 61%, depending on the scenario, while personal expenses remain unchanged, reflecting their negligible role in the adjustments. On average, the probability of being efficient in the observed dataset is 0.15, rising to 0.6, 0.67, and 0.74 in the predicted scenarios. These results suggest that some DMUs will never reach the efficiency threshold, leading the sector to converge toward a specific probability level below 1. Additionally, β values progressively increase to 1.03, 1.14, and 1.20, reflecting the growing effort required to achieve higher confidence thresholds.

While Table 5 highlights average adjustments, Table 6 focuses on median values, which offer a more representative view of typical DMU changes by reducing the impact of outliers. For the median DMU, total assets would decrease by 22% to 23%, employees by 26% to 31%, and fixed assets by 4% to 5%. Operating income would increase by 52% to 61%, consistent with the mean results. As before, personal expenses remain unaffected. The larger reductions in total assets observed in the median case indicate that smaller or more typical DMUs require relatively greater adjustments to achieve efficiency compared to larger DMUs that influence the sector average. Additionally, Table 6 reveals a significant increase in the median probability of efficiency, progressing from 0 in the observed dataset to 0.75, 0.85, and 0.95, underscoring the adaptability of the model.

Building on the observations from the mean and median values, Table 7 examines the standard deviations to highlight the variability in the projections and provide additional insights into the adjustments required. Input variables like total assets, employees, and fixed assets show reduced variability due to constraints, ensuring projections remain within reasonable limits. However, output variables such as operating income exhibit increased variability, reflecting the diversity of potential improvements among DMUs. The standard deviation of probabilities is lowest in the 0.75 scenario but increases progressively in the 0.85 and 0.95 scenarios, ultimately surpassing the variability in the observed dataset. This trend underscores the growing complexity and diversity in adjustments required to meet higher confidence thresholds.

Scenario	Observed	0.75		0.85		0.95	
Total assets	41,03	35,72	(-0,13)	35,18	(-0,14)	34,86	(-0,15)
Employees	201,00	148,29	(-0,26)	142,90	(-0,29)	139,70	(-0,30)
Fixed assets	15,28	13,51	(-0,12)	13,33	(-0,13)	13,22	(-0,13)
Personal expenses	6,76	6,75	(0,00)	6,75	(0,00)	6,75	(0,00)
Operating income	62,31	95,05	(0,53)	98,39	(0,58)	100,37	(0,61)
Beta	0,00	1,03		1,14		1,20	
Probability	0,15	0,60	(2,87)	0,67	(3,33)	0,74	(3,76)

Table 5. Mean values for projections at different confidence levels.

Scenario	Observed	0.75		0.85		0.95	
Total assets	24,56	19,12	(-0,22)	19,05	(-0,22)	18,88	(-0,23)
Employees	98,00	72,66	(-0,26)	69,87	(-0,29)	68,06	(-0,31)
Fixed assets	6,26	6,03	(-0,04)	6,01	(-0,04)	5,97	(-0,05)
Personal expenses	3,06	3,05	(0,00)	3,05	(0,00)	3,05	(0,00)
Operating income	29,14	44,35	(0,52)	45,47	(0,56)	46,90	(0,61)
Beta	0,00	0,38		0,40		0,49	
Probability	0,00	0,75	(930,79)	0,85	(1055,03)	0,95	(1179,27)

Table 6. Median values for projections at different confidence levels.

Scenario	Observed	0.75		0.85		0.95	
Total assets	49,75	45,01	(-0,10)	44,25	(-0,11)	44,14	(-0,11)
Employees	216,03	162,32	(-0,25)	157,61	(-0,27)	156,09	(-0,28)
Fixed assets	23,26	21,03	(-0,10)	20,63	(-0,11)	20,61	(-0,11)
Personal expenses	7,69	7,68	(0,00)	7,68	(0,00)	7,68	(0,00)
Operating income	84,17	115,49	(0,37)	121,50	(0,44)	122,34	(0,45)
Beta		1,65		1,90		1,91	
Probability	0,34	0,28	(-0,17)	0,32	(-0,06)	0,36	(0,05)

Table 7. Standard deviation for projections at different confidence levels.

It is observed that as the confidence level (probability threshold) increases, the required adjustments in inputs and outputs for the evaluated DMUs to be classified as efficient also increase. In other words, a higher probability threshold demands a greater magnitude of change, meaning that the evaluated firms must undergo more significant modifications to achieve efficiency. This result highlights the increasing effort required to meet stricter efficiency classification criteria.

While the aggregated results in Tables 5, 6, and 7 provide a comprehensive overview, examining specific cases can reveal additional patterns. In Table 8, we present an example of two inefficient DMUs and their projections for a scenario with a confidence level of 0.85. These DMUs were selected as representative examples to illustrate the varying adjustments required. DMU 37 fails to meet the threshold, and its best achievable performance is recorded. The adjustments required for DMU 22 are notably smaller than those for DMU 37, as shown by the percentage changes in brackets. This difference is further highlighted by the value β^*, with DMU 37's β^* being 5.55 times larger than DMU 22's, reflecting the significantly greater effort needed for DMU 37 to approach efficiency.

	DMU 22			DMU 37		
	Observed	Target		Observed	Target	
Total assets	24.71	23.20	(-0.06)	48.90	40.65	(-0.17)
Employees	212.00	196.97	(-0.07)	134.00	51.84	(-0.61)
Fixed assets	11.46	10.96	(-0.04)	20.10	17.34	(-0.14)
Personal expenses	8.00	8.00	(0.00)	6.39	6.37	(0.00)
Operating income	80.89	90.21	(0.12)	52.77	103.67	(0.96)
β^*	-	0.29		-	1.61	

Table 8. Observed values and projections for DMUs 22 and 37.

5. Conclusions and future work

A growing body of literature explores the integration of Machine Learning (ML) with Data Envelopment Analysis (DEA) to enhance efficiency analysis across various sectors. While many studies have focused on improving traditional DEA methodologies through ML techniques, our research extends this synergy by incorporating classification models to predict efficiency probabilities. This novel approach is demonstrated through an empirical analysis of SABI (Iberian Balance Sheet Analysis System) data, emphasizing its practical utility. Our findings show that the integration of ML classifiers with DEA not only predicts the efficiency status of Decision-Making Units (DMUs) but also provides a richer framework for assessing efficiency through probabilistic measures and counterfactual analysis. The advantages of our integrated approach extend beyond just analytical improvements. They also offer practical benefits in terms of scalability and adaptability. The model's ability to handle large datasets efficiently makes it especially relevant in the era of big data, where organizations across sectors are looking to exploit vast amounts of information for enhanced decision-making (Zhu, 2022). Additionally, the flexibility of the ML-DEA framework means it can be tailored to specific sector needs.

As a summary, let us point out that the new approach introduces several key methodological, interpretative, and practical contributions to efficiency analysis by integrating machine learning techniques within a DEA framework. First, we propose a novel classification-based machine learning approach in the second stage of a DEA-ML hybrid framework, moving beyond traditional regression-based techniques. In the first stage, we employ a standard DEA model to generate a binary efficiency label, which is then predicted in the second stage using classification models. Second, our framework enhances inferential power by estimating the probability of a DMU being classified as efficient, shifting DEA from a purely descriptive tool to a probabilistic efficiency assessment. This aligns efficiency analysis with modern inferential analytics and decision-making frameworks. Third, we reinterpret DEA as a classification problem, where the efficiency frontier is understood as a separating surface between technically feasible and infeasible input-output profiles, allowing efficiency measures to be framed in terms of the minimal modifications required for reclassification. Fourth, our approach is algorithm-agnostic, enabling robust efficiency assessments across various classification models, including decision trees, SVMs, neural networks, and ensemble methods. Seeking simplicity, we focus on neural networks in this paper. Fifth, we integrate Explainable AI (XAI) techniques, particularly counterfactual analysis, to define inefficiency in terms of the minimum changes required for an inefficient DMU to become efficient, offering an interpretable and actionable efficiency assessment. Sixth, we introduce a benchmarking approach that exploits the importance ranking of inputs and outputs obtained from machine learning models to assign data-driven weights to directional projections, thereby improving the interpretability and strategic value of efficiency assessments. Seventh, we enhance benchmarking by incorporating probabilistic efficiency thresholds, allowing for target setting through counterfactual benchmarking, which provides improvement strategies based on minimum necessary input-output modifications. Eighth, we propose a new ranking system for DMUs based on their probabilistic efficiency scores, offering an alternative to traditional DEA ranking methods while introducing confidence-threshold-based peer selection for more tailored benchmarking. Finally, our method facilitates a refined proximity-based benchmark identification strategy, ensuring that each DMU is compared against the closest efficient benchmark at any given efficiency probability threshold, strengthening the practical applicability of DEA for dynamic and adaptive benchmarking. These contributions collectively advance efficiency analysis by bridging the gap between DEA, statistical learning, and explainable AI, offering a more flexible and interpretable approach to performance assessment.

Looking forward, several research avenues appear promising. First. the exploration of other machine learning techniques, such as ensemble methods (e.g., Random Forest or Boosting). could provide further improvements in the robustness and accuracy of efficiency predictions. Indeed, when faced with a real empirical case, we could implement multiple machine learning techniques (and not only NN) in parallel to assess the consistency and robustness of the results. By comparing the outcomes across different models, we could evaluate the stability of efficiency classifications and ensure that our findings are not overly dependent on a specific algorithm. Secondly, the application of our integrated ML-DEA model to other domains, such as environmental sustainability and public sector performance, could be highly beneficial. These areas, where efficiency and resource optimization are critical, may significantly benefit from the enhanced analytical capabilities that our model offers. Additionally, extending our model to handle real-time data could transform operational efficiency monitoring, allowing organizations to make immediate adjustments based on current performance metrics. Lastly, further research should also focus on the development of more sophisticated counterfactual methods within the ML-DEA framework. These methods would not only enhance the interpretability of the model outcomes but also allow decision-makers to perform scenario analysis and policy testing effectively.

Acknowledgments

The authors thank the grant PID2022-136383NB-I00 funded by MCIN/AEI/ 10.13039/501100011033 and by ERDF A way of making Europe. R. Gonzalez-Moyano thanks the grant FPI/PREP2022-000117 funded by MICIU/AEI/10.13039/501100011033 and by ESF+. V. España thanks the PhD scholarship ACIF/2021/135 supported by the Conselleria d'Educació, Universitats i Ocupació (Generalitat Valenciana). Additionally, J. Aparicio thanks the grant PROMETEO/2021/063 funded by the Valencian Community (Spain).

References

Amirteimoori. A. Allahviranloo. T. Zadmirzaei. M. & Hasanzadeh. F. (2023). On the environmental performance analysis: a combined fuzzy data envelopment analysis and artificial intelligence algorithms. *Expert Systems with Applications*. *224*. 119953.

Aparicio, J., Esteve, M., Rodriguez-Sala, J. J., & Zofio, J. L. (2021). The estimation of productive efficiency through machine learning techniques: efficiency analysis trees. In *Data-enabled analytics: DEA for big data* (pp. 51-92). Cham: Springer International Publishing.

Aydin. N. & Yurdakul. G. (2020). Assessing countries' performances against COVID-19 via WSIDEA and machine learning algorithms. *Applied Soft Computing*. *97*. 106792.

Banker. R. D. & Morey. R. C. (1986). Efficiency analysis for exogenously fixed inputs and outputs. Operations Research. 34(4). 513-521.

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

Berger. A. N. Brockett. P. L. Cooper. W. W. & Pastor. J. T. (1997). New approaches for analyzing and evaluating the performance of financial institutions. *European Journal of Operational Research*. *98*(2). 170-174.

Boubaker. S. Le. T. D. Ngo. T. & Manita. R. (2023). Predicting the performance of MSMEs: A hybrid DEA-machine learning approach. *Annals of Operations Research*. 1-23.

Chambers, R. G., Chung, Y., & Färe, R. (1998). Profit, directional distance functions, and Nerlovian efficiency. Journal of optimization theory and applications, 98, 351-364.

Charles. V. Aparicio. J. & Zhu. J. (2019). The curse of dimensionality of decision-making units: A simple approach to increase the discriminatory power of data envelopment analysis. European Journal of Operational Research. 279(3). 929-940.

Charnes. A. Cooper. W. W. & Rhodes. E. (1978). Measuring the efficiency of decision making units. European Journal of Operational Research. 2(6). 429-444.

Charnes. A. Cooper. W. W Golany. B. Seiford. L. & Stutz. J. (1985). Foundations of data envelopment analysis for Pareto-Koopmans efficient empirical production functions. *Journal of econometrics*. *30*(1-2). 91-107.

Chawla. N. V. Bowyer. K. W. Hall. L. O. & Kegelmeyer. W. P. (2002). SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research. 16. 321-357.

Chen. Y. Li. Y. Xie. Q. An. Q. & Liang. L. (2014). Data envelopment analysis with missing data: a multiple imputation approach. International Journal of Information and Decision Sciences. 6(4). 315-337.

Cho. B. H. Yu. H. Lee. J. Chee. Y. J. Kim. I. Y. & Kim. S. I. (2008). Nonlinear support vector machine visualization for risk factor analysis using nomograms and localized radial basis function kernels. *IEEE Transactions on Information Technology in Biomedicine*. *12*(2). 247-256.

Cortez. P. (2010). Data mining with neural networks and support vector machines using the R/rminer tool. In *Industrial conference on data mining* (pp. 572-583). Berlin. Heidelberg: Springer Berlin Heidelberg.

Craven. M. W. & Shavlik. J. W. (1992). Visualizing learning and computation in artificial neural networks. *International journal on artificial intelligence tools*. *1*(03). 399-425.

Dadura, A. M., & Lee, T. R. (2011). Measuring the innovation ability of Taiwan's food industry using DEA. *Innovation: The European Journal of Social Science Research*, *24*(1-2), 151-172.

Daouia. A. Noh. H. & Park. B. U. (2016). Data envelope fitting with constrained polynomial splines. *Journal of the Royal Statistical Society Series B: Statistical Methodology*. *78*(1). 3-30.

Emrouznejad. A. & Shale. E. (2009). A combined neural network and DEA for measuring efficiency of large scale datasets. *Computers & Industrial Engineering*. *56*(1). 249-254.

Esteve. M. Aparicio. J. Rabasa. A. & Rodriguez-Sala. J. J. (2020). Efficiency analysis trees: A new methodology for estimating production frontiers through decision trees. Expert Systems with Applications. 162. 113783.

Esteve. M. Aparicio. J. Rodriguez-Sala. J. J. & Zhu. J. (2023). Random Forests and the measurement of super-efficiency in the context of Free Disposal Hull. European Journal of Operational Research. 304(2). 729-744.

Fallahpour. A. Olugu. E. U. Musa. S. N. Khezrimotlagh. D. & Wong. K. Y. (2016). An integrated model for green supplier selection under fuzzy environment: application of data envelopment analysis and genetic programming approach. *Neural Computing and Applications*. *27*. 707-725.

Flegl, M., Jiménez-Bandala, C. A., Sánchez-Juárez, I., & Matus, E. (2022). Analysis of production and investment efficiency in the Mexican food industry: Application of two-stage DEA. *Czech Journal of Food Sciences*, *40*(2).

Fogel. D. B. & Robinson. C. J. (2003). Techniques for extracting classification and regression rules from artificial neural networks.

Goodfellow. I. Bengio. Y. & Courville. A. (2016). Deep Learning. MIT Press.

Guerrero. N. M. Aparicio. J. & Valero-Carreras. D. (2022). Combining Data Envelopment Analysis and Machine Learning. Mathematics 2022. 10. 909.

Guillen. M. D. Aparicio. J. & España. V. J. (2023). boostingDEA: A boosting approach to Data Envelopment Analysis in R. *SoftwareX*. *24*. 101549.

Guillen. M. D. Aparicio. J. & Esteve. M. (2023). Gradient tree boosting and the estimation of production frontiers. *Expert Systems with Applications*. *214*. 119134.

Guillen. M. D. Aparicio. J. & Esteve. M. (2023). Performance Evaluation of Decision-Making Units Through Boosting Methods in the Context of Free Disposal Hull: Some Exact and Heuristic Algorithms. *International Journal of Information Technology & Decision Making*. 1-30.

Guillen. M. D. Aparicio. J. Zofío. J. L. & España. V. J. (2024). Improving the predictive accuracy of production frontier models for efficiency measurement using machine learning: The LSB-MAFS method. Computers & Operations Research. 171. 106793.

He. H. & Garcia. E. A. (2009). Learning from imbalanced data. IEEE Transactions on knowledge and data engineering. 21(9). 1263-1284.

Jin. Q. Kerstens. K. & Van de Woestyne. I. (2024). Convex and nonconvex nonparametric frontier-based classification methods for anomaly detection. OR Spectrum. 1-27.

Jomthanachai. S. Wong. W. P. & Lim. C. P. (2021). An application of data envelopment analysis and machine learning approach to risk management. *Ieee Access*. *9*. 85978-85994.

Kuhn. Max (2008). "Building Predictive Models in R Using the caret Package." *Journal of Statistical Software*. **28**(5). 1–26.

Kumar, M., & Basu, P. (2008). Perspectives of productivity growth in Indian food industry: a data envelopment analysis. *International Journal of Productivity and Performance Management*, *57*(7), 503-522.

Kuosmanen. T. & Johnson. A. L. (2010). Data envelopment analysis as nonparametric least-squares regression. *Operations Research*. *58*(1). 149-160.

Kwon. H. B. Lee. J. & Roh. J. J. (2016). Best performance modeling using complementary DEA-ANN approach: Application to Japanese electronics manufacturing firms. *Benchmarking: An International Journal*. *23*(3). 704-721.

LeCun. Y. Bengio. Y. & Hinton. G. (2015). Deep learning. Nature. 521(7553). 436-444.

Liao. Z. Dai. S. & Kuosmanen. T. (2024). Convex support vector regression. *European Journal of Operational Research*. *313*(3). 858-870.

Lin. S. W. & Lu. W. M. (2024). Using inverse DEA and machine learning algorithms to evaluate and predict suppliers' performance in the apple supply chain. *International Journal of Production Economics*. 109203.

Liu. H. H. Chen. T. Y. Chiu. Y. H. & Kuo. F. H. (2013). A comparison of three-stage DEA and artificial neural network on the operational efficiency of semi-conductor firms in Taiwan. *Modern Economy*. *4*(01). 20-31.

Lundberg. S. & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *arXiv preprint arXiv:1705.07874*.

Martens. D. Baesens. B. Van Gestel. T. & Vanthienen. J. (2007). Comprehensible credit scoring models using rule extraction from support vector machines. *European journal of operational research*. *183*(3). 1466-1476.

Nandy. A. & Singh. P. K. (2020). Farm efficiency estimation using a hybrid approach of machine-learning and data envelopment analysis: Evidence from rural eastern India. *Journal of Cleaner Production*. *267*. 122106.

Olesen. O. B. & Ruggiero. J. (2022). The hinging hyperplanes: An alternative nonparametric representation of a production function. *European Journal of Operational Research*. *296*(1). 254-266.

Olesen. O. B. Petersen. N. C. & Podinovski. V. V. (2007). Staff assessment and productivity measurement in public administration: an application of data envelopment analysis. Omega. 35(3). 297-307.

Omrani. H. Emrouznejad. A. Teplova. T. & Amini. M. (2024). Efficiency evaluation of electricity distribution companies: Integrating data envelopment analysis and machine learning for a holistic analysis. Engineering Applications of Artificial Intelligence. 133. 108636.

Parmeter. C. F. & Racine. J. S. (2013). Smooth constrained frontier analysis. *Recent Advances and Future Directions in Causality. Prediction. and Specification Analysis: Essays in Honor of Halbert L. White Jr*. 463-488.

Pastor. J. T. Lovell. C. K. & Aparicio. J. (2012). Families of linear efficiency programs based on Debreu's loss function. *Journal of Productivity Analysis*. *38*. 109-120.

Ribeiro. M. T. Singh. S. & Guestrin. C. (2016. August). " Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135-1144).

Ruck. D. W. Rogers. S. K. & Kabrisky. M. (1990). Feature selection using a multilayer perceptron. Journal of neural network computing. 2(2). 40-48.

Sexton, T. R. (1986). The methodology of data envelopment analysis. New directions for program evaluation, 32, 7-29.

Tayal. A. Solanki. A. & Singh. S. P. (2020). Integrated framework for identifying sustainable manufacturing layouts based on big data. machine learning. meta-heuristic and data envelopment analysis. *Sustainable Cities and Society*. *62*. 102383.

Thanassoulis, E., Portela, M. C., & Despic, O. (2008). Data envelopment analysis: the mathematical programming approach to efficiency analysis. The measurement of productive efficiency and productivity growth, 251-420.

Thanassoulis. E. Boussofiane. A. & Dyson. R. G. (2015). Applied data envelopment analysis. Springer.

Tickle. A. B. Andrews. R. Golea. M. & Diederich. J. (1998). The truth will come to light: Directions and challenges in extracting the knowledge embedded within trained artificial neural networks. IEEE Transactions on Neural Networks. 9(6). 1057-1068.

Tsionas. M. Parmeter. C. F. & Zelenyuk. V. (2023). Bayesian artificial neural networks for frontier efficiency analysis. *Journal of Econometrics*. *236*(2). 105491.

Tzeng. F. Y. & Ma. K. L. (2005). *Opening the black box-data driven visualization of neural networks* (pp. 383-390). IEEE.

Valero-Carreras. D. Aparicio. J. & Guerrero. N. M. (2021). Support vector frontiers: A new approach for estimating production functions through support vector machines. *Omega*. *104*. 102490.

Valero-Carreras. D. Aparicio. J. & Guerrero. N. M. (2022). Multi-output support vector frontiers. *Computers & Operations Research*. *143*. 105765.

Valero-Carreras. D. Moragues. R. Aparicio. J. & Guerrero. N. M. (2024). Evaluating different methods for ranking inputs in the context of the performance assessment of decision making units: A machine learning approach. Computers & Operations Research. 163. 106485.

Venables W.N. & Ripley B.D. (2002). *Modern Applied Statistics with S*. Fourth edition. Springer. New York. ISBN 0-387-95457-0.

Wachter. S. Mittelstadt. B. & Russell. C. (2017). ‘Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR.’ Harvard Journal of Law & Technology. 31(2). 841-887.

Wang, K., Xian, Y., Lee, C. Y., Wei, Y. M., & Huang, Z. (2019). On selecting directions for directional distance functions in a non-parametric framework: a review. *Annals of Operations Research*, *278*, 43-76.

Weiss. G. M. & Provost. F. (2003). Learning when training data are costly: The effect of class distribution on tree induction. Journal of artificial intelligence research. 19. 315-354.

Zhou. P. Ang. B. W. & Poh. K. L. (2008). A survey of data envelopment analysis in energy and environmental studies. European Journal of Operational Research. 189(1). 1-18.

Zhu. J. (2022). DEA under big data: Data enabled analytics and network data envelopment analysis. Annals of Operations Research. 309(2). 7

Zhu. N. Zhu. C. & Emrouznejad. A. (2021). A combined machine learning algorithms and DEA method for measuring and predicting the efficiency of Chinese manufacturing listed companies. *Journal of Management Science and Engineering*. *6*(4). 435-448.