



Development of a Hybrid XGBoost–Fuzzy Robust DEA Model for Evaluating the Operational and Environmental Efficiency of Petrochemical Companies Considering Undesirable Outputs and Uncertainty

Seyed Morteza Mousavi¹, Javad Gerami^{2*}, Mohammadreza Mozaffari², Roya M. Pour Ahari¹, Mohammadreza Feylizadeh²

¹ Department of Industrial Engineering, Na.C., Islamic Azad University, Najafabad, Iran

² Department of Industrial Engineering, Shi.C., Islamic Azad University, Shiraz, Iran

* Corresponding author email address: javad_gerami@iau.ac.ir

Received: 2025-06-03

Revised: 2025-10-07

Accepted: 2025-10-15

Published: 2025-11-01

Abstract

This study aimed to develop a hybrid XGBoost–Fuzzy Robust Data Envelopment Analysis model for evaluating the operational and environmental efficiency of petrochemical companies while simultaneously considering undesirable outputs, nonlinear relationships, and uncertainty in input–output data. This quantitative applied study analyzed 30 petrochemical companies over the 2020–2025 period, yielding 180 company-year observations. Inputs included total assets, number of employees, operating costs, energy consumption, and feedstock consumption, while desirable outputs comprised sales revenue, value added, and marketable production. Undesirable outputs included CO₂-equivalent emissions, air-pollutant emissions, industrial wastewater, and process or hazardous waste. A non-radial slack-based DEA model with variable returns to scale was integrated with fuzzy α -cut analysis and robust optimization to account for uncertainty. XGBoost was applied to model nonlinear efficiency determinants, with five-fold cross-validation used for hyperparameter optimization. Model performance was assessed using R², RMSE, MAE, SHAP importance, Spearman rank correlations, and sensitivity analyses across alternative model specifications. Mean efficiency declined from 0.846 under conventional DEA to 0.781 after incorporating undesirable outputs, 0.754 under fuzzy DEA, 0.727 under robust DEA, and 0.713 under the final hybrid model. The number of efficient observations decreased from 31 to 8. XGBoost achieved a test-set R² of 0.881, cross-validated R² of 0.862, RMSE of 0.047, and MAE of 0.036. Energy intensity and CO₂ emission intensity were the most influential predictors, with SHAP importance values of 0.184 and 0.162, respectively. The correlation between conventional DEA and the hybrid model was $\rho = 0.741$ ($p < 0.001$), indicating substantial ranking changes. Removing energy consumption or CO₂ emissions reduced ranking stability more than removing wastewater or hazardous waste. The hybrid XGBoost–Fuzzy Robust DEA model provided a more discriminating and uncertainty-sensitive assessment of petrochemical efficiency than conventional DEA, demonstrating that operational performance should be evaluated jointly with environmental burdens, nonlinear determinants, and data uncertainty.

Keywords: XGBoost; Fuzzy DEA; Robust DEA; Operational Efficiency; Environmental Efficiency; Petrochemical Companies; Undesirable Outputs; Uncertainty

How to cite this article:

Mousavi, S. M., Gerami, J., Mozaffari, M., Pour Ahari, R. M., & Feylizadeh, M. (2025). Development of a Hybrid XGBoost–Fuzzy Robust DEA Model for Evaluating the Operational and Environmental Efficiency of Petrochemical Companies Considering Undesirable Outputs and Uncertainty. *Management Strategies and Engineering Sciences*, 7(6), 1-17.

1. Introduction

The petrochemical industry occupies a strategic position in modern industrial economies because it supplies intermediate and final products to a wide range of sectors,

including plastics, construction, transportation, agriculture, pharmaceuticals, textiles, packaging, and advanced manufacturing. At the same time, petrochemical production is characterized by high capital intensity, substantial energy use, complex material flows, significant dependence on



fossil-based feedstocks, and considerable environmental externalities. These characteristics make performance evaluation in petrochemical companies substantially more complex than conventional productivity assessment based solely on financial outputs or production volume. In particular, a petrochemical company may appear operationally efficient when evaluated according to output maximization and cost control while simultaneously generating excessive greenhouse-gas emissions, wastewater, hazardous residues, or other undesirable outputs. The contemporary sustainability agenda therefore requires efficiency measurement systems to account simultaneously for economic performance, resource consumption, environmental burdens, technological heterogeneity, and uncertainty. The importance of this integrated perspective is reinforced by evidence that the global chemical and petrochemical sector faces increasing pressure to move toward low- and zero-emission production trajectories while maintaining industrial competitiveness and continuity of supply [1]. More recent analyses of petrochemical carbon emissions have also shown that the sector remains a major contributor to industrial CO₂ emissions and that identifying high-impact production units and emission sources is essential for carbon-neutrality planning [2]. Accordingly, operational efficiency can no longer be interpreted independently of environmental efficiency.

The environmental dimension of petrochemical performance extends well beyond carbon emissions. Petrochemical plants generate diverse forms of gaseous, liquid, and solid pollutants, and their environmental management requires substantial technological and organizational resources. Global research on petrochemical waste gases has demonstrated the sustained growth of scientific and technological interest in emission control, reflecting the increasing importance of air-pollution mitigation in this industry [3]. Wastewater treatment is another critical dimension because petrochemical effluents frequently contain refractory organic compounds, toxic contaminants, and chemically complex mixtures that cannot always be efficiently treated using conventional processes. Research on hybrid biological reactors has identified several technical factors that constrain petrochemical wastewater-treatment efficiency, including pollutant loading, hydraulic conditions, and treatment configuration [4]. Advanced oxidation processes have therefore been investigated as complementary technologies for improving the degradation of persistent contaminants in petrochemical wastewater [5], while more recent studies have explored sustainable

ozonation based on natural zeolite catalysts as a pathway toward improved treatment performance and reduced environmental burden [6]. Wastewater reuse through electrodeionization has likewise been proposed as an important strategy for reducing freshwater demand and promoting circular water management within petrochemical plants [7]. In parallel, recovery and valorization of useful chemicals from petrochemical wastewater indicate that environmental management may generate additional economic value rather than functioning solely as a compliance cost [8]. Solid hazardous residues create a further challenge, as demonstrated by studies on mechanochemical treatment and dechlorination of petrochemical sludge using industrial-waste additives [9]. These findings collectively demonstrate that environmentally responsible petrochemical performance involves multiple undesirable outputs that must be incorporated explicitly into efficiency assessment rather than treated as peripheral indicators.

Energy efficiency is equally central to petrochemical competitiveness and sustainability. Petrochemical operations involve highly energy-intensive processes, including cracking, separation, distillation, compression, heating, cooling, and feedstock transformation. Improvements in process integration and cogeneration can therefore substantially affect both economic and environmental outcomes. Simulation studies of ethylene plants have shown that alternative cogeneration configurations can improve overall energy efficiency and reduce resource losses [10]. Intelligent diagnostic methods have also been developed to identify energy-saving potential in petrochemical plants, highlighting the usefulness of machine-learning and support-vector approaches for recognizing nonlinear energy-performance patterns [11]. At the macro-industrial level, shifts in energy and transportation systems are expected to alter petrochemical demand structures, production technologies, and competitive conditions, thereby making conventional static efficiency assessment increasingly inadequate [12]. In the Iranian context, the transition toward cleaner energy sources has particular relevance because petrochemical production constitutes an important source of industrial value added, while the substitution of green electricity may contribute to renewable-energy development and lower-carbon production structures [13]. These developments indicate that petrochemical efficiency must be evaluated through frameworks capable of capturing the interaction between

energy intensity, production output, environmental emissions, and technological change.

Operational efficiency within petrochemical companies also depends on labor productivity, capital utilization, process reliability, production scheduling, and investment quality. The effect of investment on labor productivity in chemical and petrochemical industries suggests that capital allocation and technological modernization can materially influence productive performance [14]. Similarly, industrial scheduling represents a complex optimization problem because petrochemical systems include interconnected units, capacity constraints, maintenance requirements, inventory interactions, and uncertain operating conditions; systematic evidence from industrial scheduling research emphasizes the need for advanced optimization approaches that can handle such complexity [15]. Network-level performance is also important because petrochemical clusters operate as interconnected production ecosystems rather than isolated plants. Analysis of cluster performance and network properties has shown that structural interactions among firms, infrastructure, material flows, and production linkages can affect overall system performance [16]. Strategic development of petrochemical industries therefore depends not only on individual plant efficiency but also on broader priorities related to modernization, investment, integration, and competitiveness [17]. These considerations strengthen the case for efficiency models that can discriminate among companies with heterogeneous scales, technologies, resources, and operating environments.

Digitalization and Industry 4.0 have further transformed the way petrochemical performance can be measured, predicted, and optimized. Industry 4.0 practices combined with lean manufacturing have been associated with improvements in sustainability performance in petrochemical and plastic organizations, indicating that digital integration can strengthen operational, environmental, and organizational outcomes simultaneously [18]. Digital twin technologies have increasingly been applied to safety and operational improvement in petrochemical environments by enabling virtual representation, process monitoring, ergonomic analysis, and proactive risk control [19]. Digitalization has also extended to integrated simulation environments, as illustrated by online object-oriented platforms developed for modeling petrochemical plant networks and supporting complex plant-level analysis [20]. Drones represent another emerging component of industrial digitalization because they can support inspection, monitoring, maintenance, and data

acquisition in complex manufacturing environments, although their implementation also introduces operational and regulatory challenges [21]. These developments suggest that contemporary efficiency assessment should benefit from machine-learning methods capable of exploiting nonlinear patterns in large and heterogeneous industrial datasets.

Machine learning is particularly relevant because relationships among petrochemical inputs, outputs, emissions, and operational conditions are rarely linear. Artificial-intelligence-based controllers and nonlinear control systems have increasingly been proposed for petrochemical process applications because they can handle dynamic interactions, nonlinearities, and uncertain process behavior more effectively than conventional linear approaches [22]. Machine-learning models have also been successfully used to predict controlled variables in advanced process-control systems using petrochemical time-series data, demonstrating the value of data-driven methods for identifying complex operational dependencies [23]. In addition, modern superstructure-based optimization models have been developed to improve the design and production efficiency of emerging petrochemical and bio-based materials, illustrating how optimization and predictive analytics can jointly support process decision-making [24]. The broader transition beyond conventional petrochemicals toward industrial-scale biomanufacturing further indicates that future process systems will involve increasingly complex combinations of biological, chemical, and computational technologies [25]. In this context, ensemble learning methods such as Extreme Gradient Boosting (XGBoost) offer important analytical advantages because they can model nonlinear associations, capture interactions, rank variable importance, tolerate heterogeneous predictors, and generate robust out-of-sample predictions. However, machine learning by itself does not construct an efficiency frontier or determine how far each production unit is from best-practice performance.

Data Envelopment Analysis (DEA) provides a complementary methodological perspective because it evaluates the relative efficiency of decision-making units by comparing multiple inputs and outputs without requiring a predefined production function. DEA is particularly appropriate for petrochemical companies because these firms simultaneously consume capital, labor, feedstock, and energy while generating financial and physical outputs. Previous research has applied DEA to assess eco-efficiency in petrochemical enterprises, showing that the method can

identify substantial differences in performance across companies and can incorporate environmental considerations into benchmarking [26]. Research on industrial agglomeration and green technological innovation has similarly demonstrated that eco-efficiency in petrochemical regions is influenced by structural and technological factors rather than by production scale alone [27]. Nevertheless, conventional DEA frequently relies on deterministic observations and may therefore produce unstable or overly optimistic scores when the underlying data contain reporting errors, measurement variability, environmental uncertainty, or temporal fluctuations. This limitation is particularly important in petrochemical settings because emissions, energy use, wastewater production, operating costs, and capacity utilization can fluctuate considerably even within the same production unit.

The treatment of undesirable outputs constitutes another methodological challenge. Traditional DEA models are generally structured around the assumption that outputs should be increased, but pollutant emissions, wastewater, industrial waste, and safety incidents are outputs that should be reduced rather than expanded. Excluding these variables can lead to an incomplete interpretation of efficiency because firms may be classified as efficient simply because they maximize production regardless of environmental consequences. The need for integrated environmental evaluation is consistent with strategic environmental-assessment principles, which emphasize the incorporation of environmental and health impacts into industrial decision-making [28]. Environmental consequences also have direct implications for human health, as petrochemical exposure has been associated with toxicological and epigenetic risks, further increasing the societal importance of reducing pollution at the source [29]. Petrochemical sustainability is also closely connected with operational safety, since accident patterns within industrial complexes provide important information about systemic vulnerabilities and long-term sustainability performance [30]. Risk-management studies in petrochemical engineering projects similarly indicate that operational efficiency cannot be separated from risk control, because process failures and unmanaged hazards generate economic losses, environmental damage, and disruption of productive capacity [31]. Even seemingly specific engineering issues such as corrosion in desalination systems can affect plant reliability, maintenance requirements, and resource consumption, thereby influencing broader efficiency outcomes [32]. Comparative studies of gas-leak detection

systems have likewise shown the importance of probabilistic and simulation-based approaches for improving safety performance in process industries [33].

Uncertainty therefore represents a fundamental element of petrochemical performance measurement. Production systems operate under fluctuating feedstock characteristics, energy prices, maintenance conditions, market demand, environmental-monitoring error, equipment degradation, and stochastic operating disturbances. Treating all recorded input and output values as perfectly precise can generate rankings that are highly sensitive to minor data changes. Fuzzy DEA provides a useful approach for representing imprecise values through membership functions and α -cut intervals, while robust DEA protects efficiency estimates against plausible deviations from nominal observations. Combining fuzzy and robust approaches can therefore improve the reliability of frontier estimation by distinguishing genuinely efficient firms from those whose rankings depend strongly on uncertain measurements. Such methodological integration is increasingly relevant in an industry undergoing technological and structural transformation. For example, carbon capture has been examined as an additional decarbonization option for petrochemical and other carbon-intensive sectors, but its contribution depends on technological configuration, energy requirements, and sector-specific conditions [34]. Similarly, the development of petrochemical systems must increasingly be evaluated not only in terms of incremental efficiency improvement but also within wider debates regarding industrial transformation, resource dependence, and the future scale and composition of synthetic-material production [35]. These transitions imply that efficiency models should remain robust across alternative environmental and technological scenarios.

At the same time, a strictly frontier-based methodology may not fully exploit the predictive information contained in complex industrial data. XGBoost can capture nonlinear relationships between operational characteristics and efficiency outcomes, identify the relative importance of variables such as energy intensity, emission intensity, capacity utilization, and feedstock conversion, and detect interaction effects that conventional deterministic models may overlook. Integrating XGBoost with Fuzzy Robust DEA can therefore provide a more comprehensive analytical architecture: DEA supplies a nonparametric efficiency frontier, fuzzy logic represents imprecision, robust optimization protects the frontier against uncertainty, and XGBoost provides nonlinear predictive learning and feature-

importance information. Such a hybrid framework is particularly suitable for petrochemical companies because performance emerges from the interaction of production, finance, energy, environmental management, safety, and technological factors. Moreover, the increasingly sophisticated digital and computational environment of petrochemical operations creates a practical basis for integrating optimization with machine learning rather than treating them as separate analytical paradigms.

Despite the increasing sophistication of petrochemical sustainability research, important methodological gaps remain. Existing studies frequently focus on one dimension at a time, such as energy optimization, emission reduction, wastewater treatment, process control, safety, or conventional eco-efficiency. For example, system-dynamics modeling has been used to explore pathways for emission reduction in petrochemical industries [36], while green technology innovation has been examined in relation to regional eco-efficiency [27]; however, these approaches do not directly combine firm-level frontier efficiency with predictive machine learning and explicit uncertainty modeling. Likewise, environmental-treatment research provides valuable technical evidence on pollution reduction but does not necessarily determine whether a company remains operationally efficient after accounting for undesirable outputs. The strategic transition of the sector also requires flexible evaluation methods because future petrochemical systems may incorporate circularity, renewable energy, carbon capture, advanced digital control, and bio-based production simultaneously. Consequently, there is a clear need for an integrated framework capable of assessing heterogeneous petrochemical companies while simultaneously considering desirable outputs, undesirable environmental outputs, nonlinear relationships, and uncertainty in industrial data.

Therefore, the aim of this study was to develop a hybrid XGBoost–Fuzzy Robust DEA model for evaluating the operational and environmental efficiency of petrochemical companies by simultaneously incorporating desirable outputs, undesirable environmental outputs, nonlinear performance relationships, and uncertainty in input–output data.

2. Methodology

This study employed a quantitative, applied, and analytical research design to develop and evaluate a hybrid XGBoost–Fuzzy Robust Data Envelopment Analysis (DEA)

model for simultaneously assessing the operational and environmental efficiency of petrochemical companies under conditions of uncertainty and in the presence of undesirable outputs. The units of analysis consisted of petrochemical companies regarded as decision-making units (DMUs), rather than individual human participants. The statistical population included petrochemical companies whose financial, operational, energy-consumption, production, and environmental information was accessible through corporate reports, audited financial statements, stock-market disclosures, sustainability and environmental reports, and records available from relevant industrial institutions in Tehran. Following an initial screening of the population, 30 petrochemical companies were included in the final analysis. Companies were selected using a purposive census-oriented approach based on predetermined eligibility criteria, including continuous operation during the study period, availability of comparable financial and production information, availability of data required for measuring environmental performance, consistency of accounting periods, and absence of substantial missing information for the variables incorporated into the efficiency model. Companies with incomplete observations for major input, desirable-output, or undesirable-output variables were excluded to prevent biased efficiency estimates. Panel data covering six consecutive fiscal years from 2020 to 2025 were collected for the 30 companies, yielding a balanced analytical dataset of 180 company-year observations. The longitudinal structure of the dataset was selected to reduce sensitivity to temporary production shocks and to allow the proposed hybrid model to capture heterogeneity and temporal variations in company performance. The conceptual framework considered operational efficiency and environmental efficiency as interconnected dimensions of overall organizational performance. Accordingly, the DEA component was designed to identify the relative efficiency frontier, whereas the XGBoost component was employed to model nonlinear relationships among operational, financial, energy, and environmental variables and to improve the predictive and discriminatory capacity of the efficiency assessment. Uncertainty in observed inputs and outputs, particularly energy use, operating costs, production quantities, and pollutant emissions, was incorporated through fuzzy and robust optimization procedures.

Data were collected using a researcher-developed data extraction checklist specifically designed in accordance with the structure of the proposed hybrid XGBoost–Fuzzy Robust DEA framework. The checklist was used to systematically

extract quantitative information from audited annual financial statements, board of directors' reports, official company disclosures, production statistics, energy-consumption reports, environmental monitoring records, and sustainability-related documentation. Before the final extraction stage, the checklist was reviewed in terms of content relevance, operational definition, comparability, and data availability to ensure that the selected variables were theoretically appropriate for efficiency analysis and empirically measurable across the participating companies. The variables were classified into three major groups: inputs, desirable outputs, and undesirable outputs. The principal input variables consisted of total assets as a proxy for capital employed, number of employees representing labor input, total operating costs, total energy consumption expressed in standardized energy-equivalent units, and feedstock or raw-material consumption. These indicators were selected because petrochemical production is simultaneously capital-intensive, energy-intensive, and dependent on substantial intermediate material consumption. Desirable outputs comprised annual sales revenue, value added, and total marketable petrochemical production, thereby capturing both the physical and economic dimensions of productive performance. Undesirable outputs included carbon dioxide-equivalent greenhouse-gas emissions, total air-pollutant emissions where comparable data were available, industrial wastewater generation, and hazardous or process-related waste generation. Treating environmental pollutants as undesirable outputs made it possible to distinguish companies that achieved high production levels through environmentally costly processes from companies that generated similar levels of economic output with lower environmental burdens.

To operationalize uncertainty, the extracted values were not assumed to be perfectly deterministic. Variables susceptible to measurement error, reporting variation, energy-price fluctuations, production instability, and environmental monitoring uncertainty were represented by fuzzy numbers. Triangular fuzzy numbers were used to define lower, most plausible, and upper values for uncertain observations. The central value corresponded to the reported or observed value, whereas the lower and upper bounds were determined using observed interannual variation, dispersion in the corresponding company-level data, and predefined uncertainty intervals. This specification enabled the DEA model to reflect realistic variations in the data rather than treating each recorded value as an error-free quantity. An alpha-cut transformation was subsequently used to convert

fuzzy variables into interval representations at different confidence levels, allowing efficiency estimates to be examined under alternative degrees of uncertainty. The robust component of the model additionally incorporated uncertainty sets around selected input and output coefficients so that the resulting efficiency scores would remain stable when data deviated moderately from their nominal values.

A machine-learning dataset was constructed from the same standardized indicators for implementation of the XGBoost algorithm. In addition to the principal DEA variables, a set of supplementary explanatory variables was incorporated to improve prediction of efficiency, including asset turnover, operating-cost intensity, energy intensity per unit of production, labor productivity, feedstock conversion efficiency, emission intensity, wastewater intensity, waste intensity, return on assets, and production-capacity utilization. Continuous variables exhibiting substantial scale differences were transformed or standardized where necessary, and extreme observations were examined using percentile distributions and interquartile-range criteria. Outliers representing plausible industrial conditions were retained because extreme energy use, emissions, or production efficiency may contain meaningful information concerning best-practice and inefficient DMUs. Only observations attributable to identifiable reporting errors were corrected using the original corporate documentation. Missing values for core DEA variables were not statistically imputed because artificial estimation of inputs or outputs could alter the position of the efficiency frontier; therefore, companies with insufficient core data were excluded during the eligibility assessment.

Data analysis was conducted through a multistage hybrid procedure integrating XGBoost, fuzzy logic, robust optimization, and Data Envelopment Analysis. Initially, descriptive statistics, including the mean, standard deviation, minimum, maximum, median, coefficient of variation, and annual growth rate, were calculated for all inputs, desirable outputs, undesirable outputs, and supplementary variables. Correlation analysis was used as an exploratory procedure to examine the direction and magnitude of relationships among candidate indicators and to identify variables exhibiting excessive redundancy. Because DEA does not require distributional assumptions, normality was not considered a prerequisite for inclusion of variables. Nevertheless, distributions were inspected to identify severe skewness, extraordinary values, and inconsistencies in measurement scales. Input and output variables were subsequently

normalized where required for integration with the machine-learning component, whereas the original economically interpretable values were retained for the DEA optimization models.

The first analytical component consisted of training an Extreme Gradient Boosting model to identify complex nonlinear relationships between organizational characteristics and efficiency-related performance. The dataset was partitioned into training and testing subsets using an 80:20 ratio, while five-fold cross-validation was applied within the training set to optimize model hyperparameters and reduce overfitting. The principal hyperparameters subjected to tuning included the number of boosting trees, maximum tree depth, learning rate, minimum child weight, subsampling ratio, column-sampling ratio, and L1 and L2 regularization parameters. Predictive performance was evaluated using the coefficient of determination, root mean square error, and mean absolute error. Feature importance was assessed using gain-based measures and SHAP values to determine the relative contribution of operational, financial, energy, and environmental indicators to variations in predicted efficiency. The XGBoost results were not used as a replacement for DEA efficiency scores; rather, the algorithm served as a complementary learning mechanism for detecting nonlinear importance structures, estimating expected performance patterns, and generating data-driven information that was subsequently incorporated into the hybrid efficiency framework.

The DEA component was specified as a non-radial, slack-based efficiency model capable of incorporating undesirable outputs directly into the production technology. Each company-year observation was treated as an individual DMU in the longitudinal efficiency assessment. Both operational resources and environmental burdens were incorporated simultaneously so that a company could attain a high overall efficiency score only when it generated comparatively high desirable output while using fewer resources and producing lower levels of pollution. Desirable outputs were modeled according to the principle that larger quantities were preferred, whereas undesirable outputs were modeled according to the principle that reductions represented improvements in environmental performance. Variable returns to scale were adopted because petrochemical companies differ substantially in production capacity, asset size, technological configuration, feedstock access, and market position, making the assumption of constant returns to scale unnecessarily restrictive. The model generated overall efficiency scores as well as input excesses,

desirable-output shortfalls, and undesirable-output excesses, thereby providing specific improvement targets for inefficient companies.

Fuzzy DEA was subsequently implemented to account for imprecision in selected observed values. Triangular fuzzy numbers were transformed using alpha-cut levels ranging from 0.1 to 0.9, and lower-bound and upper-bound efficiency scores were estimated for each DMU at the corresponding uncertainty levels. This procedure produced efficiency intervals rather than a single deterministic score and therefore allowed the stability of company rankings to be assessed as uncertainty increased. The robust optimization component was then integrated into the DEA formulation by introducing bounded uncertainty sets around uncertain coefficients. Different robustness-budget parameters were examined to control the trade-off between protection against data uncertainty and conservativeness of the efficiency estimates. A company whose efficiency remained comparatively high across alternative alpha levels and robustness parameters was interpreted as having structurally stable performance, whereas substantial deterioration in efficiency under modest uncertainty indicated that the deterministic efficiency estimate was sensitive to data variation.

In the final hybrid stage, the information generated through XGBoost was integrated with the Fuzzy Robust DEA framework. Variables identified by XGBoost as having negligible explanatory relevance were examined for potential redundancy, while the relative predictive contribution and nonlinear interaction structure identified by the machine-learning model were used as auxiliary information in the construction and interpretation of the final efficiency model. DEA efficiency scores obtained under deterministic, fuzzy, and fuzzy-robust conditions were compared with XGBoost-predicted efficiency values to evaluate consistency between frontier-based and machine-learning-based assessments. Spearman rank correlations were calculated to assess the stability of company rankings across alternative model specifications. Differences between deterministic DEA and the proposed hybrid model were also examined to determine whether explicit consideration of undesirable outputs and uncertainty materially changed efficiency classifications.

Model validity and robustness were evaluated through several sensitivity procedures. Efficiency scores were recalculated after systematic modification of uncertainty levels, robust-budget parameters, selected input-output combinations, and XGBoost hyperparameter specifications.

Leave-one-variable-out analyses were conducted to determine whether the overall ranking structure was excessively dependent on any single input or output. Furthermore, conventional DEA results were compared with Fuzzy DEA, Robust DEA, and the proposed XGBoost–Fuzzy Robust DEA model to assess the incremental analytical contribution of each methodological component. A DMU was regarded as technically efficient when its efficiency score reached the relevant frontier value under the specified model, while companies below the frontier were evaluated according to their distance from best-practice performance and their corresponding input, output, and environmental slacks. All statistical preprocessing and XGBoost analyses were conducted in Python using appropriate machine-learning libraries, whereas mathematical programming formulations for the fuzzy and robust DEA models were solved using optimization software capable of linear and robust optimization. Statistical significance for inferential procedures was evaluated at the 0.05 level, and all analyses were performed using reproducible computational scripts to ensure consistency of the analytical process.

3. Findings and Results

The final analytical sample consisted of 30 petrochemical companies and 180 company-year observations covering the six-year period from 2020 to 2025. In terms of company size,

11 companies (36.7%) were classified as large-scale petrochemical producers, 12 companies (40.0%) as medium-scale producers, and 7 companies (23.3%) as relatively smaller production units according to their average total assets and nominal production capacity during the study period. With respect to the principal production orientation, 9 companies (30.0%) were predominantly engaged in polymer and downstream petrochemical production, 8 companies (26.7%) were primarily involved in olefin-related production, 7 companies (23.3%) were mainly active in aromatic and related petrochemical products, and 6 companies (20.0%) had a diversified product portfolio. The mean operating history of the companies was 20.63 years ($SD = 7.84$), with a range from 8 to 39 years. The average nominal capacity utilization across the study period was 78.42% ($SD = 9.61$), while the annual company-level utilization rates ranged from 51.73% to 96.84%. Of the 180 company-year observations, 87 (48.3%) belonged to companies with average annual production exceeding 1.5 million tons, whereas 93 observations (51.7%) were associated with companies producing below this threshold. Environmental reporting completeness increased over the study period, with standardized reporting of greenhouse-gas emissions available for all included observations, while comparable wastewater and industrial-waste data were verified from corporate and environmental records prior to inclusion in the analytical dataset.

Table 1. Descriptive Statistics of Input, Desirable-Output, and Undesirable-Output Variables

Variable	Mean	SD	Minimum	Median	Maximum	CV (%)
Total assets, trillion IRR	186.42	101.76	42.81	161.35	512.64	54.59
Number of employees	1,684.37	812.45	486.00	1,507.00	4,395.00	48.23
Operating costs, trillion IRR	74.58	42.19	15.47	65.62	219.83	56.57
Energy consumption, thousand TOE	913.26	488.31	194.52	814.68	2,426.57	53.47
Feedstock consumption, thousand tons	1,782.64	1,012.73	324.16	1,536.82	5,114.73	56.81
Sales revenue, trillion IRR	134.57	82.46	24.36	111.92	421.77	61.28
Value added, trillion IRR	48.31	28.74	8.93	40.75	151.62	59.49
Marketable production, thousand tons	1,467.82	872.56	276.94	1,221.37	4,568.24	59.45
CO ₂ -equivalent emissions, thousand tons	1,218.53	691.42	248.37	1,041.66	3,782.18	56.74
Air-pollutant emissions, tons	3,764.28	2,286.17	612.41	3,174.55	11,824.63	60.73
Industrial wastewater, thousand m ³	2,086.47	1,181.95	397.82	1,759.64	6,273.91	56.65
Process and hazardous waste, tons	7,463.18	4,581.26	1,184.37	6,219.83	23,548.72	61.38

As presented in Table 1, substantial heterogeneity was observed among the petrochemical companies in terms of production scale, resource utilization, economic output, and environmental burden. Total assets averaged 186.42 trillion IRR, while the coefficient of variation of 54.59% indicated considerable dispersion across company-year observations.

Similar variability was observed for operating costs, energy consumption, and feedstock consumption, with coefficients of variation exceeding 53%. Marketable production averaged 1,467.82 thousand tons per year but ranged from 276.94 to 4,568.24 thousand tons, confirming the presence of substantial scale differences among the DMUs and

supporting the use of a variable-returns-to-scale DEA specification. Sales revenue and value added also exhibited pronounced cross-sectional variability, with coefficients of variation of 61.28% and 59.49%, respectively. Regarding undesirable outputs, mean CO₂-equivalent emissions were 1,218.53 thousand tons per company-year, while mean industrial wastewater generation was 2,086.47 thousand m³. Air-pollutant emissions and process or hazardous waste

exhibited coefficients of variation above 60%, suggesting that environmental burdens varied markedly even among units with comparable production levels. Taken together, these descriptive findings indicated that the sample contained sufficient variation in inputs, desirable outputs, and undesirable outputs to discriminate between relatively efficient and inefficient companies through the proposed frontier-based model.

Table 2. Operational and Environmental Efficiency Scores Under Alternative DEA Specifications

Efficiency Measure	Mean	SD	Minimum	Median	Maximum	Efficient Observations, n (%)
Conventional operational DEA	0.846	0.108	0.521	0.867	1.000	31 (17.2)
Environmental DEA with undesirable outputs	0.781	0.126	0.413	0.798	1.000	21 (11.7)
Fuzzy DEA, $\alpha = 0.5$	0.754	0.129	0.391	0.768	0.992	14 (7.8)
Robust DEA	0.727	0.132	0.362	0.741	0.976	10 (5.6)
XGBoost–Fuzzy Robust DEA	0.713	0.128	0.348	0.724	0.968	8 (4.4)

Table 2 demonstrates a systematic decline in estimated efficiency as progressively more realistic production and environmental constraints were incorporated into the analytical framework. Under the conventional operational DEA model, which primarily emphasized the conversion of productive inputs into desirable outputs, the mean efficiency score was 0.846 (SD = 0.108), and 31 of the 180 company-year observations (17.2%) were positioned on the efficiency frontier. Once undesirable environmental outputs were explicitly incorporated, mean efficiency declined to 0.781, while the number of efficient observations decreased to 21 (11.7%). This reduction indicated that several companies that appeared operationally efficient generated disproportionately high environmental burdens relative to their economic and physical output. Incorporation of fuzzy

uncertainty at $\alpha = 0.5$ further reduced mean efficiency to 0.754, demonstrating that part of the apparent deterministic efficiency was sensitive to uncertainty in the underlying operational and environmental data. The robust DEA specification produced an even lower mean score of 0.727, and only 10 observations remained technically efficient. Finally, the complete XGBoost–Fuzzy Robust DEA model generated a mean efficiency score of 0.713 (SD = 0.128), with values ranging from 0.348 to 0.968. Only 8 observations, corresponding to 4.4% of the analytical sample, remained close to the robust best-practice frontier. The consistent contraction of the efficient set showed that conventional DEA tended to provide comparatively optimistic efficiency estimates when environmental externalities and data uncertainty were omitted.

Table 3. XGBoost Predictive Performance and Relative Importance of Efficiency Determinants

Indicator	Value
Training-set R ²	0.934
Test-set R ²	0.881
Five-fold cross-validated R ²	0.862
Test RMSE	0.047
Test MAE	0.036
Energy intensity, SHAP importance	0.184
CO ₂ emission intensity, SHAP importance	0.162
Capacity utilization, SHAP importance	0.137
Feedstock conversion efficiency, SHAP importance	0.119
Operating-cost intensity, SHAP importance	0.103
Labor productivity, SHAP importance	0.087
Wastewater intensity, SHAP importance	0.076
Asset turnover, SHAP importance	0.058
Hazardous-waste intensity, SHAP importance	0.043
Return on assets, SHAP importance	0.031

The XGBoost model demonstrated strong predictive performance in estimating efficiency-related patterns from the operational, financial, energy, and environmental indicators. The model explained 93.4% of the variance in the training sample and 88.1% of the variance in the test sample, while the mean five-fold cross-validated coefficient of determination was 0.862. The relatively limited difference between training and testing performance suggested that the regularization and cross-validation procedures effectively restricted excessive overfitting. The test RMSE was 0.047 and the MAE was 0.036, indicating relatively small prediction errors given that efficiency scores were bounded between zero and one. SHAP-based feature importance analysis showed that energy intensity was the most

influential variable, accounting for an importance value of 0.184, followed by CO₂ emission intensity at 0.162 and production-capacity utilization at 0.137. Feedstock conversion efficiency and operating-cost intensity were also important predictors, with SHAP importance values of 0.119 and 0.103, respectively. Labor productivity, wastewater intensity, and asset turnover exhibited moderate predictive importance, whereas hazardous-waste intensity and return on assets contributed comparatively less to prediction. These findings indicated that the operational-environmental efficiency of the petrochemical companies was determined more strongly by resource conversion, energy requirements, emissions, and productive utilization than by conventional profitability indicators alone.

Table 4. Efficiency Estimates Across Fuzzy α -Cut Levels in the Proposed Model

α -Cut Level	Mean Lower-Bound Efficiency	Mean Central Efficiency	Mean Upper-Bound Efficiency	Mean Interval Width
0.1	0.654	0.713	0.779	0.125
0.3	0.669	0.713	0.762	0.093
0.5	0.682	0.713	0.747	0.065
0.7	0.694	0.713	0.734	0.040
0.9	0.706	0.713	0.721	0.015

As shown in Table 4, the uncertainty intervals surrounding the hybrid efficiency scores narrowed systematically as the α -cut level increased. At $\alpha = 0.1$, representing the broadest uncertainty range, mean lower-bound efficiency was 0.654 and mean upper-bound efficiency was 0.779, producing an average interval width of 0.125. At $\alpha = 0.5$, the corresponding lower and upper efficiency bounds were 0.682 and 0.747, respectively, and the mean interval width decreased to 0.065. At the highest evaluated α -cut level of 0.9, the interval width narrowed to only 0.015, with lower- and upper-bound estimates of 0.706 and 0.721 around the central efficiency score of 0.713. The

central estimate remained unchanged across α -cut levels because it represented the most plausible or nominal configuration of the fuzzy observations, whereas the bounds captured the range of possible efficiency outcomes attributable to data uncertainty. The progressive narrowing of the intervals confirmed the internal consistency of the fuzzy formulation and also showed that several company-year efficiency estimates were materially sensitive to uncertainty at lower α -cut levels. Consequently, the use of a single deterministic efficiency estimate would conceal meaningful information regarding the reliability and stability of company-level performance assessments.

Table 5. Comparison of Model Rankings, Rank Stability, and Robustness Results

Comparison / Sensitivity Test	Spearman ρ	Mean Absolute Rank Change	Maximum Rank Change	p-value
Conventional DEA vs. environmental DEA	0.846	12.37	46	<0.001
Environmental DEA vs. fuzzy DEA	0.921	7.84	31	<0.001
Fuzzy DEA vs. robust DEA	0.947	5.93	24	<0.001
Robust DEA vs. XGBoost–Fuzzy Robust DEA	0.962	4.28	19	<0.001
Conventional DEA vs. XGBoost–Fuzzy Robust DEA	0.741	17.62	58	<0.001
Hybrid model after removing energy consumption	0.883	9.46	33	<0.001
Hybrid model after removing CO ₂ emissions	0.901	8.11	29	<0.001
Hybrid model after removing wastewater	0.968	3.72	17	<0.001
Hybrid model after removing hazardous waste	0.976	3.15	14	<0.001

The ranking comparison presented in Table 5 showed that the relative ordering of petrochemical companies was

progressively modified as environmental factors, fuzzy uncertainty, robust optimization, and XGBoost-derived

nonlinear information were incorporated. The Spearman correlation between conventional DEA and environmental DEA rankings was 0.846, indicating strong but incomplete agreement, with a mean absolute rank change of 12.37 positions and a maximum displacement of 46 positions. The association became stronger between conceptually adjacent models: environmental DEA and fuzzy DEA yielded a correlation of 0.921, fuzzy DEA and robust DEA yielded 0.947, and robust DEA and the final hybrid model yielded 0.962. However, the correlation between conventional DEA and the complete XGBoost–Fuzzy Robust DEA ranking was only 0.741, and the mean absolute rank change reached 17.62 positions, with some observations changing by as many as 58 positions. This result demonstrated that

incorporation of undesirable outputs and uncertainty was sufficiently influential to alter managerial interpretations of relative performance. Sensitivity analysis further showed that removing energy consumption from the hybrid model reduced rank stability to $\rho = 0.883$, while removing CO₂ emissions produced $\rho = 0.901$. By contrast, omission of wastewater or hazardous waste caused smaller ranking changes, with correlations of 0.968 and 0.976, respectively. These results were consistent with the XGBoost importance analysis and confirmed that energy consumption and greenhouse-gas emissions were particularly influential in defining the efficiency frontier and differentiating company performance.

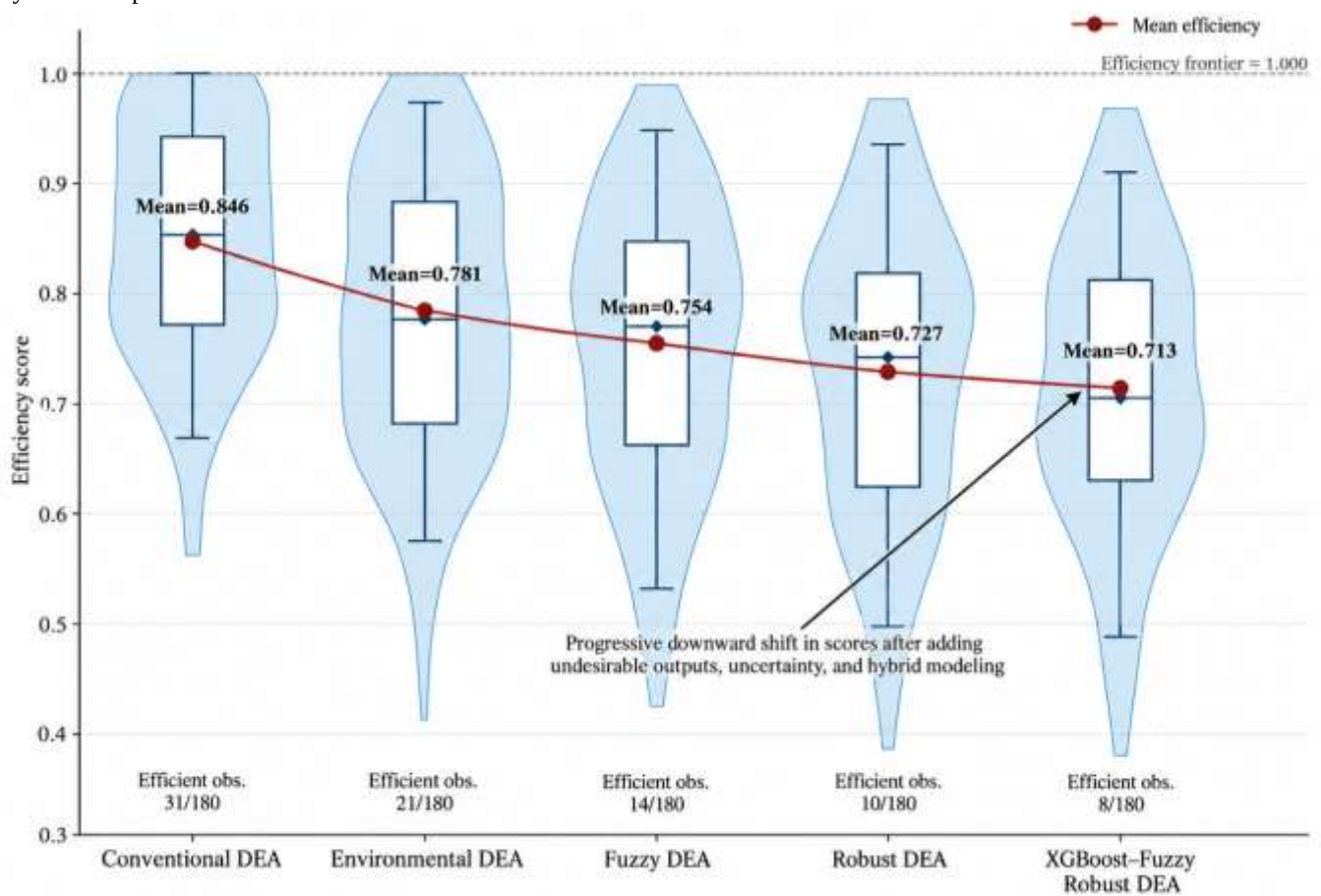


Figure 1. Distribution and Comparative Ranking of Petrochemical Company Efficiency Scores Across Conventional DEA, Environmental DEA, Fuzzy DEA, Robust DEA, and the Proposed XGBoost–Fuzzy Robust DEA Model

The comparative distribution illustrated in Figure 1 showed a visible downward shift in efficiency scores from the conventional DEA model toward the proposed hybrid specification. Conventional DEA produced the most concentrated distribution near the upper end of the efficiency scale, whereas inclusion of undesirable outputs increased

dispersion and shifted a considerable number of observations toward intermediate efficiency values. The fuzzy and robust specifications further differentiated companies whose apparent performance depended heavily on nominal data assumptions. Under the final XGBoost–Fuzzy Robust DEA model, the number of observations

positioned near the efficiency frontier was substantially reduced, while the separation between high-, medium-, and low-performing DMUs became more pronounced. Company-year observations characterized by high capacity utilization, relatively low energy consumption per unit of output, efficient feedstock conversion, and comparatively low CO₂ intensity generally remained in the upper portion of the final ranking. In contrast, several companies that performed favorably under conventional DEA experienced marked downward ranking adjustments when their high emission intensities, energy requirements, wastewater generation, or sensitivity to uncertainty were incorporated. This pattern confirmed that the proposed model provided a more discriminating representation of operational and environmental efficiency than a production-oriented DEA model based exclusively on deterministic desirable inputs and outputs.

4. Discussion and Conclusion

The present study developed and evaluated a hybrid XGBoost–Fuzzy Robust DEA framework for simultaneous assessment of operational and environmental efficiency in petrochemical companies, and the findings showed that the estimated efficiency of the studied firms declined progressively as the analytical model moved from conventional DEA toward environmental, fuzzy, robust, and fully hybrid specifications. The mean efficiency score decreased from 0.846 under conventional operational DEA to 0.781 after incorporating undesirable outputs, 0.754 under fuzzy DEA, 0.727 under robust DEA, and 0.713 under the final XGBoost–Fuzzy Robust DEA model. At the same time, the number of observations classified as efficient fell from 31 company-year observations under conventional DEA to only 8 under the final hybrid model. This pattern indicates that a substantial proportion of the apparent efficiency observed under conventional analysis was conditional on ignoring environmental burdens and uncertainty. In other words, some petrochemical companies that appeared efficient in converting capital, labor, energy, and feedstock into production and revenue were no longer positioned near the best-practice frontier once greenhouse-gas emissions, wastewater, waste generation, and data imprecision were considered. This finding is consistent with previous DEA-based research demonstrating that eco-efficiency assessment in petrochemical enterprises can yield materially different conclusions from purely production-oriented efficiency analysis [26]. It also corresponds with evidence that

petrochemical eco-efficiency is shaped by green technological innovation and structural characteristics, rather than production growth alone [27]. The observed downward adjustment is therefore theoretically meaningful because it distinguishes productive performance from genuinely sustainable productive performance.

The reduction in efficiency following incorporation of undesirable outputs is particularly important in the context of the petrochemical industry because production expansion is commonly accompanied by considerable environmental externalities. In the present study, the inclusion of CO₂-equivalent emissions, air pollutants, industrial wastewater, and process or hazardous waste reduced both average efficiency and the number of frontier observations. This result suggests that firms with high marketable output or sales were not necessarily environmentally efficient if their production depended on disproportionate pollution generation. This interpretation aligns with the growing body of evidence emphasizing the carbon intensity of petrochemical production and the need to identify high-impact contributors to emissions [2]. It is also consistent with the broader zero-emission transition literature, which shows that the chemical and petrochemical sector cannot achieve long-term sustainability through output expansion alone and requires substantial improvements in energy systems, process technologies, and emission-control strategies [1]. The result also accords with system-dynamics evidence indicating that petrochemical emission reduction requires structural and technological interventions rather than isolated operational adjustments [36]. From this perspective, the environmental DEA results demonstrate that treating pollution as an explicit undesirable output provides a more realistic efficiency frontier because it penalizes production configurations that externalize environmental costs.

The descriptive and model-based findings further showed that energy intensity and CO₂ emission intensity were the two most influential variables in the XGBoost component, with SHAP importance values of 0.184 and 0.162, respectively. Capacity utilization, feedstock conversion efficiency, and operating-cost intensity were also highly influential. These results indicate that operational and environmental efficiency in petrochemical production is fundamentally driven by the quality with which firms convert energy and raw materials into saleable products while limiting emissions. The dominant role of energy intensity is in line with previous simulation-based research showing that improvement of cogeneration and energy-

integration systems can significantly enhance the energy efficiency of ethylene petrochemical plants [10]. It also supports intelligent energy-diagnosis research demonstrating that advanced computational methods can identify energy-saving opportunities in complex petrochemical operations [11]. The importance of energy-related variables is additionally consistent with studies examining broader transitions in energy and transportation systems, which suggest that petrochemical competitiveness will increasingly depend on adaptation to lower-carbon energy structures [12]. In the Iranian context, the finding is particularly relevant because substitution of conventional electricity with greener energy sources has been proposed as a pathway for simultaneously supporting renewable-energy development and petrochemical value creation [13]. Thus, the present model confirms that energy performance is not a secondary technical indicator but a central determinant of relative efficiency.

The strong role of CO₂ emission intensity also reinforces the need to evaluate petrochemical companies through integrated economic-environmental criteria. A company may possess high production capacity and favorable financial performance yet remain comparatively inefficient if each unit of output is associated with high greenhouse-gas emissions. This finding is compatible with current analyses of petrochemical carbon-neutrality pathways, which emphasize both plant-level emission heterogeneity and the importance of identifying high-impact contributors [2]. It is also consistent with research on carbon capture applications, which suggests that petrochemical decarbonization requires sector-specific technological interventions and that emission mitigation cannot be evaluated independently of process configuration and energy demand [34]. The environmental interpretation of the results is strengthened by research showing that industrial sustainability requires formal assessment of environmental and health consequences in addition to economic performance [28]. The health dimension is particularly relevant because petrochemical exposure has been associated with adverse toxicological and epigenetic outcomes, implying that reducing emissions provides benefits extending beyond corporate compliance and directly into public-health protection [29]. Therefore, the lower efficiency of emission-intensive companies in the proposed hybrid framework represents an analytically appropriate adjustment rather than an artificial penalty.

The results concerning wastewater and hazardous waste were also noteworthy. Although wastewater intensity and hazardous-waste intensity had lower feature importance than

energy and CO₂ variables, both contributed to differentiation among firms, and omission of these variables produced measurable changes in ranking. This finding is consistent with extensive technical literature showing that petrochemical wastewater is difficult to treat because of its chemically complex and often refractory pollutant composition. Previous work on hybrid biological reactors has documented efficiency-limiting factors in petrochemical wastewater treatment [4], while studies of advanced oxidation processes have demonstrated the need for more intensive treatment technologies to address persistent contaminants [5]. More recent work on sustainable ozonation using zeolite-based catalysts similarly indicates that improved treatment efficiency can reduce the environmental burden associated with petrochemical effluents [6]. The present findings are also aligned with research on wastewater reuse through electrodeionization, which demonstrates that treated effluent can potentially be reintegrated into petrochemical operations and thereby improve both water efficiency and sustainability [7]. The relatively smaller but still meaningful role of wastewater and waste variables in the hybrid model suggests that environmental efficiency is multidimensional and that no single pollutant indicator is sufficient to characterize a plant's environmental performance.

The importance of waste-related indicators is further supported by research on resource recovery and hazardous-residue treatment. Previous studies have shown that valuable chemical compounds may be recovered from petrochemical wastewater and subsequently reused, illustrating the potential for transforming an environmental liability into a productive resource [8]. Similarly, mechanochemical approaches to dechlorination of petrochemical sludge demonstrate that hazardous solid residues can be managed using innovative treatment strategies [9]. These findings support the interpretation that firms achieving lower waste intensity may be benefiting from superior process control, waste minimization, recycling, recovery, or treatment practices. Consequently, including waste indicators in DEA improves the ability of the model to distinguish companies that achieve similar levels of production through different environmental pathways. This is particularly important in industries where physical output is often comparable but environmental management practices differ substantially across companies.

A major methodological contribution of the present study was the explicit incorporation of fuzzy uncertainty and robust optimization. The results showed that efficiency

intervals widened at lower α -cut levels and narrowed progressively as the α level increased, while central efficiency remained stable at 0.713. This pattern indicates that the underlying data contain meaningful uncertainty but that the central ordering of firms was not arbitrary. The progressive decline in average efficiency from fuzzy DEA to robust DEA further suggests that nominal observations may overstate efficiency when data are assumed to be perfectly precise. This result is plausible in petrochemical environments, where energy consumption, emissions, production volumes, and operating costs can vary due to maintenance schedules, feedstock characteristics, market fluctuations, unplanned shutdowns, and monitoring differences. The importance of uncertainty is also consistent with petrochemical risk-management research showing that process performance is exposed to multiple operational and engineering risks [31]. Similarly, studies of accident data in petrochemical industrial complexes demonstrate that operational sustainability is affected by variability in system conditions and safety performance [30]. The present results therefore support the use of uncertainty-aware efficiency models, particularly when managerial decisions may depend on relatively small differences in rankings.

The XGBoost component exhibited strong predictive performance, with a test-set R^2 of 0.881, cross-validated R^2 of 0.862, RMSE of 0.047, and MAE of 0.036. The limited difference between training and test performance indicates that the model captured stable nonlinear patterns rather than merely fitting noise in the training data. This finding is consistent with prior applications of machine learning in petrochemical process control and prediction. Machine-learning-based prediction of controlled variables in advanced process-control systems has demonstrated that nonlinear models can extract useful relationships from complex petrochemical time-series data [23]. Similarly, broader reviews of artificial-intelligence and nonlinear control approaches show that such techniques are increasingly relevant for petrochemical systems characterized by dynamic, interconnected, and nonlinear process behavior [22]. The present findings extend this literature by using machine learning not as a replacement for frontier analysis, but as a complementary mechanism for identifying important variables and nonlinear performance structures. This hybridization is also consistent with current advances in process optimization, where superstructure-based models are used to improve the design and operation of complex chemical production systems [24].

The results concerning capacity utilization and feedstock conversion efficiency further emphasize the operational side of petrochemical performance. Companies with higher utilization of productive capacity and more effective transformation of feedstock tended to occupy more favorable positions in the hybrid model, provided that these gains were not achieved at the cost of excessive emissions or energy consumption. This finding is consistent with studies demonstrating that investment and technological modernization can improve labor productivity and industrial performance in chemical and petrochemical sectors [14]. It also aligns with industrial scheduling literature showing that better coordination of production resources and constraints can improve system performance [15]. At the cluster level, network properties and coordination among interconnected petrochemical facilities have likewise been shown to affect performance [16]. These prior findings support the interpretation that operational efficiency emerges from coordinated utilization of capacity, labor, feedstock, technology, and infrastructure rather than from any single input.

The strong ranking changes observed between conventional DEA and the final hybrid model are another important result. Although adjacent models produced high rank correlations, the correlation between conventional DEA and XGBoost–Fuzzy Robust DEA declined to 0.741, with a mean absolute rank displacement of 17.62 positions and a maximum change of 58 positions. This demonstrates that the choice of efficiency model has substantive managerial consequences. A company classified as a strong performer under deterministic operational DEA may be reclassified when environmental burdens and uncertainty are taken into account. The sensitivity analysis reinforced this interpretation: removing energy consumption or CO₂ emissions produced larger changes in ranking than removing wastewater or hazardous waste. This pattern is consistent with the feature-importance results and demonstrates that the hybrid framework is not simply generating different rankings due to methodological complexity; rather, the ranking changes are systematically associated with variables that are central to petrochemical sustainability. This is also compatible with the broader literature on Industry 4.0 and lean manufacturing, which suggests that integrated technological and organizational improvements can simultaneously affect operational and sustainability performance [18]. Digital twin technologies offer another relevant parallel because they integrate multiple dimensions of operational data to improve safety and process

performance rather than evaluating variables in isolation [19].

The findings also have implications for the broader technological transformation of the petrochemical sector. Digital platforms and modeling environments increasingly make it possible to integrate plant-level data across interconnected units, supporting more sophisticated evaluation and optimization [20]. Emerging manufacturing technologies such as drones can further expand the collection of inspection and operational data, which may enhance future efficiency models [21]. At the same time, the industry is entering a period in which conventional petrochemical production may increasingly coexist with bio-based and alternative manufacturing systems. Current research on industrial-scale biomanufacturing highlights both opportunities and technical challenges in moving beyond conventional petrochemical pathways [25]. This broader transition strengthens the value of flexible efficiency frameworks that can accommodate heterogeneous technologies, environmental impacts, and uncertain inputs. The critical literature on petrochemical transformation similarly emphasizes that sectoral sustainability involves not only technological substitution but also deeper reconsideration of the structure and scale of material production [35]. Thus, the proposed hybrid framework may be particularly useful in comparative settings where firms differ substantially in technological maturity and environmental strategy.

Overall, the study indicates that the XGBoost–Fuzzy Robust DEA framework provides a more conservative, discriminating, and multidimensional assessment of petrochemical efficiency than conventional DEA. The model simultaneously recognizes operational productivity, environmental responsibility, nonlinear performance determinants, and uncertainty in observed data. This interpretation is consistent with the strategic development literature, which emphasizes that petrochemical competitiveness increasingly depends on modernization, technological capability, sustainability, and improved resource utilization [17]. It is also compatible with research on industrial sustainability and safety, in which accident prevention, process reliability, environmental performance, and operational efficiency are viewed as interrelated dimensions rather than independent objectives [30]. Moreover, specific engineering issues such as corrosion control in desalination and process-water systems can affect maintenance burden and operational continuity, indirectly influencing efficiency [32], while improved leak-detection

technologies can reduce losses and environmental risks in process industries [33]. Collectively, these observations support the central conclusion that petrochemical efficiency should be evaluated through an integrated framework capable of differentiating firms that achieve high production through resource-efficient and environmentally responsible processes from those whose apparent performance depends on excessive energy consumption, emissions, or environmentally costly production structures.

A principal limitation of the present study concerns the scope and structure of the available data. The analysis was restricted to 30 petrochemical companies and 180 company-year observations over the 2020–2025 period, and although this panel was sufficient for the proposed modeling framework, the results may not fully represent all petrochemical production configurations, particularly smaller private firms, highly specialized plants, or companies with substantially different technologies. Environmental variables were derived from available corporate and industrial records, and differences in measurement protocols, reporting frequency, or monitoring accuracy may have introduced residual heterogeneity despite the use of fuzzy and robust methods. In addition, some potentially relevant environmental indicators, such as volatile organic compounds, water stress, biodiversity impacts, detailed chemical oxygen demand, and lifecycle emissions, could not be incorporated consistently across all DMUs. The XGBoost component was designed to identify predictive relationships rather than causal effects, and therefore the estimated feature importance should not be interpreted as evidence that changing a particular variable would necessarily cause a proportional change in efficiency. Finally, the proposed framework inevitably depends on decisions regarding input-output selection, fuzzy uncertainty bounds, robustness parameters, and DEA specification, and alternative methodological choices could generate somewhat different efficiency estimates.

Future research should extend the proposed framework across larger multi-country petrochemical samples and longer time periods to examine whether the identified performance patterns remain stable across institutional, technological, and regulatory contexts. Comparative analysis between oil- and gas-based petrochemical systems, integrated refinery-petrochemical complexes, polymer producers, and emerging bio-based production facilities would provide a more detailed understanding of technological heterogeneity. Future studies could also incorporate dynamic or network DEA to model internal

production stages and intertemporal productivity change, as well as Malmquist or global productivity indices to distinguish technological progress from efficiency catch-up. The uncertainty component could be expanded through probabilistic, distributionally robust, stochastic, or type-2 fuzzy formulations and compared directly with the present triangular fuzzy approach. On the machine-learning side, future research could compare XGBoost with random forests, LightGBM, CatBoost, neural networks, support vector regression, and ensemble stacking, while using SHAP interaction values and partial-dependence analysis to investigate nonlinear relationships in greater detail. Integrating lifecycle environmental indicators, carbon pricing scenarios, renewable-energy penetration, circularity measures, and plant-level digitalization variables would further improve the comprehensiveness of future efficiency assessments.

For practice, petrochemical managers should use efficiency benchmarking as a multidimensional decision-support process rather than relying exclusively on production volume, profitability, or conventional technical-efficiency scores. Particular managerial attention should be directed toward energy intensity, greenhouse-gas emission intensity, capacity utilization, feedstock conversion, and operating-cost control, because these variables exerted the strongest influence on hybrid efficiency. Firms with favorable production performance but weak environmental scores should prioritize energy integration, process optimization, waste-heat recovery, cleaner power procurement, leak reduction, emission monitoring, and wastewater reuse rather than attempting to improve rankings through output expansion alone. Companies should also institutionalize integrated operational and environmental databases so that energy, production, emissions, waste, maintenance, and financial data can be analyzed jointly and at regular intervals. The proposed hybrid model can be used as a benchmarking dashboard to identify stable frontier firms, detect units whose rankings are sensitive to uncertainty, and establish realistic improvement targets for inefficient DMUs. Regulators and industry associations can likewise use such models to design performance-based incentives, differentiated environmental targets, and support programs that recognize both productive efficiency and pollution reduction.

Authors' Contributions

Authors equally contributed to this article.

Acknowledgments

Authors thank all participants who participate in this study.

Declaration of Interest

The authors report no conflict of interest.

Funding

According to the authors, this article has no financial support.

Ethical Considerations

All procedures performed in this study were under the ethical standards.

References

- [1] D. Saygin and D. Gielen, "Zero-Emission Pathway for the Global Chemical and Petrochemical Sector," *Energies*, vol. 14, no. 13, p. 3772, 2021, doi: 10.3390/en14133772.
- [2] T. Lei, W. Zhao, S. Ma, C. Lai, and D. Guan, "China's Petrochemical Plants' CO₂ Emissions and High-Impact Contributors for Carbon-Neutrality Production," *Science Advances*, vol. 12, no. 12, 2026, doi: 10.1126/sciadv.adx7784.
- [3] M. Wu *et al.*, "Global Trends in the Research and Development of Petrochemical Waste Gas From 1981 to 2022," *Sustainability*, vol. 16, no. 14, p. 5972, 2024, doi: 10.3390/su16145972.
- [4] M. E. Ahmed, A. Mydlarczyk, and A. Al-Haddad, "Efficiency Limiting Factors of Petrochemical Wastewater Treatment Using Hybrid Biological Reactor," *Journal of Environmental Engineering and Landscape Management*, vol. 30, no. 3, pp. 380-392, 2022, doi: 10.3846/jeelm.2022.17633.
- [5] S. Verma, P. Kumar, V. C. Srivastava, and U. L. Štangar, "Application of Advanced Oxidation Processes (AOPs) for the Treatment of Petrochemical Industry Wastewater," pp. 103-128, 2021, doi: 10.1007/978-3-030-83811-9_6.
- [6] Y. Yuan *et al.*, "Sustainable Ozonation Using Natural Zeolite-Based Catalysts for Petrochemical Wastewater Treatment," *Sustainability*, vol. 17, no. 20, p. 9110, 2025, doi: 10.3390/su17209110.
- [7] A. B. d. Santos, A. Giacobbo, M. A. S. Rodrigues, and A. M. Bernardes, "Electrodeionization for Wastewater Reuse in Petrochemical Plants," *Water*, vol. 16, no. 3, p. 401, 2024, doi: 10.3390/w16030401.
- [8] J. Hernández-Fernández, R. Ortega-Toro, and J. López, "A New Route of Valorization of Petrochemical Wastewater: Recovery of 1,3,5-Tris (4-Tert-Butyl-3-Hydroxy-2,6-Dimethyl Benzyl)-1,3,5-Triazine-2,4,6-(1h,3h,5h)-Trione (Cyanox 1790) and Its Subsequent Application in a PP Matrix to Improve Its Thermal Stability," *Molecules*, vol. 28, no. 5, p. 2003, 2023, doi: 10.3390/molecules28052003.
- [9] N. Alavi, S. Adabi, M. Sadani, A. Eslami, and M. M. Amini, "Mechanochemical Dechlorination of Petrochemical Sludge Through a Planetary Ball Mill and Using Industrial Wastes as Additives," *Environmental Progress & Sustainable Energy*, vol. 41, no. 4, 2022, doi: 10.1002/ep.13828.

- [10] L. F. d. Silva, G. Martins, A. G. Gallego, and R. Palacios-Bereche, "Simulation Studies of the Existing Cogeneration System and Alternative Conceptions to Improve the Energy Efficiency for an Ethylene Petrochemical Plant," *International Journal of Energy Research*, vol. 46, no. 14, pp. 20728-20746, 2022, doi: 10.1002/er.8546.
- [11] B. Zhao, D. Qin, D. Gao, and L. Xu, "Energy Saving Diagnosis Model of Petrochemical Plant Based on Intelligent Curvelet Support Vector Machine," *Soft Computing*, vol. 25, no. 24, pp. 15391-15401, 2021, doi: 10.1007/s00500-021-06151-z.
- [12] I. Giannikopoulos, A. Skouteris, T. F. Edgar, M. Bâldea, D. T. Allen, and M. A. Stadther, "Probing the Impact of an Energy and Transportation Paradigm Shift on the Petrochemicals Industry," *Industrial & Engineering Chemistry Research*, vol. 61, no. 33, pp. 12169-12179, 2022, doi: 10.1021/acs.iecr.2c00309.
- [13] M. K. Aghababaei, A. Hajinezhad, S. F. Moosavian, and R. Fattahi, "Evaluating the Substitution Value of Green Electricity for Advancing Renewable Energy Development Through the Value-Added Contributions of Iran's Petrochemical Industries: A System Dynamics Approach," *Energy Science & Engineering*, vol. 14, no. 2, pp. 793-807, 2025, doi: 10.1002/ese3.70382.
- [14] V. Ismayilov, N. Gafarov, and S. Valiyeva, "Impact of Investment on Labour Productivity in Azerbaijan's Chemical and Petrochemical Industry," *Qubahan Academic Journal*, vol. 4, no. 2, pp. 198-213, 2024, doi: 10.48161/qaj.v4n2a211.
- [15] H. E. Boukrouma, Y. Zennir, and R. Bendib, "A Systematic Literature Review on Industrial Scheduling Optimization," *Algerian Journal of Signals and Systems*, vol. 9, no. 4, pp. 222-235, 2024, doi: 10.51485/ajss.v9i4.241.
- [16] M. D. Tan, P. Ibarra-González, I. Nikolić, and A. Ramírez, "Determining the Performance and Network Properties of Petrochemical Clusters," pp. 1211-1216, 2023, doi: 10.1016/b978-0-443-15274-0.50193-1.
- [17] U. Shalbolova, L. Salykova, K. Bissenov, and S. Makhanov, "Priorities for the Development of the Petrochemical Industry as Part of the Fuel and Energy Complex of Kazakhstan," *E3s Web of Conferences*, vol. 411, p. 01043, 2023, doi: 10.1051/e3sconf/202341101043.
- [18] A. M. Ghaithan, M. A. Khan, A. Mohammed, and L. A. Hadidi, "Impact of Industry 4.0 and Lean Manufacturing on the Sustainability Performance of Plastic and Petrochemical Organizations in Saudi Arabia," *Sustainability*, vol. 13, no. 20, p. 11252, 2021, doi: 10.3390/su132011252.
- [19] J. Jia *et al.*, "Digital Twin Technology and Ergonomics for Comprehensive Improvement of Safety in the Petrochemical Industry," *Process Safety Progress*, vol. 43, no. 3, pp. 507-522, 2024, doi: 10.1002/prs.12575.
- [20] M. Fakhroleslam, "Vira Workspace: A Novel Online Object-Oriented Framework as a SaaS for the Modeling and Simulation of Petrochemical Plant Networks," *Acs Omega*, vol. 10, no. 37, pp. 42202-42217, 2025, doi: 10.1021/acsomega.5c00403.
- [21] O. Maghazei, "Drones in Manufacturing: Opportunities and Challenges," pp. 290-298, 2022, doi: 10.46254/eu05.20220073.
- [22] T. S. Ansari and S. A. A. Taqvi, "State-of-the-Art Review on the Applications of Nonlinear and Artificial Intelligence-Based Controllers in Petrochemical Processes," *Chembioeng Reviews*, vol. 10, no. 6, pp. 884-906, 2023, doi: 10.1002/cben.202300017.
- [23] J. H. Lee, Y. Kim, I. J. Kim, S. B. Hong, and H. S. Yun, "Comparative Analysis of Ultrasonic and Traditional Gas-Leak Detection Systems in the Process Industries: A Monte Carlo Approach," *Processes*, vol. 12, no. 1, p. 67, 2023, doi: 10.3390/pr12010067.
- [24] Y. Chen, "A Superstructure-based Optimization Model for Polyethylene Furandicarboxylate Production," *Aiche Journal*, 2025, doi: 10.1002/aic.70079.
- [25] J. H. Kim, H. E. Yu, M. H. Kim, and S. Y. Lee, "Beyond Petrochemicals: Challenges and Opportunities in Industrial-Scale Biomanufacturing," *Nature Communications*, vol. 17, no. 1, 2026, doi: 10.1038/s41467-026-73835-1.
- [26] Y. Wang, L. Yao, S. Cui, and Z. Zhu, "Eco-Efficiency Assessment of Chinese Petrochemical Enterprises: A Data Envelopment Analysis Approach," *Energy & Environment*, vol. 33, no. 6, pp. 1160-1180, 2021, doi: 10.1177/0958305x211065424.
- [27] M. Cui and X. Wu, "Effects of Petrochemical Industry Agglomeration And Green Technology Innovation On Eco-Efficiency: An Analysis Based On Provincial Data in China," *Polish Journal of Environmental Studies*, vol. 33, no. 2, pp. 1043-1055, 2023, doi: 10.15244/pjoes/172761.
- [28] V. G. Iyer, "Importance of Strategic Environmental Assessment (Sea) Process and Environmental Health Impact Assessment (Ehia) Process Towards Sustainable Climate Change and Control," *Journal of Intelligent Computing and Mathematics*, 2024, doi: 10.55571/jicm.2024.01028.
- [29] S. Jayaraman, A. Eswaran, V. P. Veeraraghavan, M. A. Fazal, A. Al-Rahbi, and S. R. Sirasanagandla, "Implications of Petrochemical Exposure and Epigenetic Alterations on Human Health," *Frontiers in Toxicology*, vol. 7, 2025, doi: 10.3389/ftox.2025.1542871.
- [30] I. s. Kim, C. Yoon, D. Lee, G. A. Shin, and S. Jung, "A Study on the Sustainability of Petrochemical Industrial Complexes Through Accident Data Analysis," *Processes*, vol. 12, no. 12, p. 2637, 2024, doi: 10.3390/pr12122637.
- [31] J. P. B. Visbal, R. Ortega-Toro, and J. Hernández-Fernández, "Application of Risk Management in Applied Engineering Projects in a Petrochemical Plant Producing Polyvinyl Chloride in Cartagena, Colombia," *Chemengineering*, vol. 9, no. 4, p. 75, 2025, doi: 10.3390/chemengineering9040075.
- [32] H. K. Thabet, O. A. A. El-Shamy, A. M. Ashmawy, and M. A. Deyab, "The Impact of Corrosion Inhibitors in Desalination Systems," *Acs Omega*, vol. 8, no. 48, pp. 45224-45231, 2023, doi: 10.1021/acsomega.3c07129.
- [33] M. Lee *et al.*, "Machine Learning-Based Prediction of Controlled Variables of APC Systems Using Time-Series Data in the Petrochemical Industry," *Processes*, vol. 11, no. 7, p. 2091, 2023, doi: 10.3390/pr11072091.
- [34] Q. Jiang, "Comparison of Additional Empowerment of Carbon Capture Applications in Petrochemical, Construction, Power Industry," *E3s Web of Conferences*, vol. 424, p. 03005, 2023, doi: 10.1051/e3sconf/202342403005.
- [35] J. P. Tilsted, "Degrowing a Synthetic World: A Critical Realist Perspective on Petrochemical Transformation," *Environmental Values*, vol. 35, no. 3, pp. 204-225, 2025, doi: 10.1177/09632719251357480.
- [36] E. Sylvia and Y. Sunitiyoso, "A System Dynamic Model for Indonesian Petrochemical Industry Emission Reduction," *International Journal of Energy Sector Management*, vol. 18, no. 1, pp. 1-25, 2023, doi: 10.1108/ijesm-07-2022-0006.