Jurnal Teknik Informatika C.I.T Medicom, 18 (2) (2026) 79-94

Published by: Institute of Computer Science (IOCS)

Jurnal Teknik Informatika C.I.T Medicom

Journal homepage: www.medikom.iocspublisher.org

Machine Learning Integration in DEA Models: Current Developments and Future Challenges

Hengki Tamando Sihotang[1], Fristi Riandari[2], Rasenda[3], Wildan Alrasyid[4]

[1] Sains Data, Universitas Pembangunan Nasional Veteran Jakarta

[2] Politeknik Negeri Medan, Medan, Indonesia

[3,4] Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta

Article Info

Article history

Received : March 27, 2026
Revised : April 26, 2026
Accepted : May 18, 2026

Keywords:

Data Envelopment Analysis (DEA);
Machine Learning (ML);
Efficiency Analysis;
Artificial Intelligence (AI);
Decision Support Systems (DSS).

Abstract

The increasing availability of large and complex datasets has created new opportunities for enhancing Data Envelopment Analysis (DEA) through the integration of Machine Learning (ML) techniques. This study reviews current developments in the integration of ML and DEA models and identifies key challenges, trends, and future research opportunities. A systematic literature review was conducted by examining recent studies that combine DEA with various machine learning algorithms across multiple application domains, including healthcare, banking and finance, manufacturing, supply chain management, energy, agriculture, and higher education. The findings indicate that Artificial Neural Networks (ANN), Support Vector Machines (SVM), Random Forests, Gradient Boosting methods, and Deep Learning models are among the most frequently employed techniques in DEA-ML frameworks. Despite these advantages, several challenges remain, including data quality issues, model interpretability, computational complexity, limited generalizability, and the lack of standardized integration frameworks. The review concludes that the integration of ML and DEA offers substantial potential for advancing efficiency analysis and organizational performance evaluation. Future research should focus on developing explainable artificial intelligence (XAI) solutions, real-time efficiency analytics, federated learning approaches, and standardized hybrid DEA-ML frameworks to improve transparency, scalability, and practical applicability across diverse operational environments.

Corresponding Author:

Hengki Tamando Sihotang
Sains Data, Universitas Pembangunan Nasional Veteran Jakarta
Jl. Rs. Fatmawati, Pondok Labu Kota Jakarta Selatan 12450 DKI Jakarta
Email: hengkisihotang@upnvj.ac.id

1. Introduction

Efficiency measurement has become an essential aspect of organizational and industrial performance evaluation in an increasingly competitive and data-driven environment. Organizations across sectors such as healthcare, manufacturing, banking, energy, education, and logistics continuously seek methods to optimize resource utilization while maximizing outputs(Bughin et al., 2017). Accurate efficiency assessment enables decision-makers to identify best practices, allocate resources effectively, improve productivity, and maintain sustainable competitive advantages. As organizations generate vast amounts of operational data, traditional performance evaluation methods often struggle to capture the complexity of modern production and service systems. Consequently, advanced analytical techniques have become increasingly important for evaluating organizational efficiency.

One of the most widely used approaches for efficiency assessment is Data Envelopment Analysis (DEA), a non-parametric mathematical programming technique introduced to evaluate the relative efficiency of decision-making units (DMUs) that utilize multiple inputs to produce multiple outputs. DEA has gained significant attention due to its ability to measure efficiency without requiring predefined functional relationships between inputs and outputs(Cooper et al., 2007). Unlike traditional parametric methods, DEA constructs an efficiency frontier from observed data and compares each DMU against the best-performing units within the sample. This flexibility has made DEA a valuable tool in numerous application domains, including healthcare systems, educational institutions, financial organizations, transportation networks, and energy management.

Despite its advantages, DEA faces several challenges in contemporary data environments. Traditional DEA models rely heavily on predefined input-output variables and assume that the available data adequately represent the production process. In practice, however, organizations increasingly operate in environments characterized by high-dimensional, heterogeneous, dynamic, and large-scale datasets. Furthermore, DEA models may be sensitive to noise, outliers, and variable selection issues, potentially affecting the accuracy and robustness of efficiency evaluations. As data complexity continues to increase, there is a growing need for complementary analytical approaches capable of overcoming these limitations.

At the same time, Machine Learning (ML) has emerged as a transformative field within artificial intelligence, offering powerful tools for data analysis, prediction, pattern recognition, classification, clustering, and decision support. Machine learning algorithms can process large volumes of structured and unstructured data, uncover nonlinear relationships, and automatically identify hidden patterns that may not be apparent through conventional statistical techniques. Advances in supervised, unsupervised, and deep learning methods have enabled organizations to extract valuable insights from complex datasets and support more informed decision-making processes. Consequently, machine learning has been increasingly adopted across various sectors to improve operational efficiency, forecasting accuracy, and strategic planning.

The integration of Machine Learning and Data Envelopment Analysis has recently attracted considerable attention from researchers and practitioners(Jomthanachai et al., 2021). By combining the strengths of DEA in efficiency measurement with the predictive and analytical capabilities of ML, hybrid frameworks can address many limitations associated with traditional DEA models. Machine learning techniques can assist DEA through feature selection, dimensionality reduction, clustering of decision-making units, outlier detection, efficiency prediction, and post-efficiency analysis. In turn, DEA can provide valuable efficiency indicators that serve as inputs for machine learning models, creating synergistic approaches for performance evaluation and optimization. This integration has opened new opportunities for developing more robust, scalable, and intelligent efficiency assessment frameworks.

The integration of Machine Learning (ML) and Data Envelopment Analysis (DEA) has emerged as a rapidly growing research area over the past decade. One of the early challenges identified in DEA research was the selection of appropriate input and output variables(Nataraja & Johnson, 2011). To address this issue, Benítez-Peña, Bogetoft, and Romero Morales (2020) proposed a mathematical optimization framework for feature selection in DEA. Their study demonstrated that incorporating feature selection mechanisms could improve model robustness and reduce the impact of irrelevant variables on efficiency estimation. The authors argued that intelligent variable selection represents a critical step toward integrating data-driven learning techniques into DEA-based analyses.

As machine learning techniques became increasingly popular, researchers began investigating how ML algorithms could complement DEA in performance prediction and classification tasks. Shi and Zhao (2023) developed an integrated machine learning and DEA framework for performance outcome prediction in high-dimensional and imbalanced datasets. Their approach utilized machine learning techniques to replicate DEA-based efficiency classifications and improve predictive performance. The study highlighted the potential of ML algorithms to enhance the interpretability and robustness of DEA applications, particularly in complex decision-making environments characterized by large numbers of variables.

During the same period, researchers also explored the integration of clustering and decision-support techniques with DEA. A comprehensive review conducted by Gomes, Osiro, and Lima (2023) examined the combination of DEA, Multi-Criteria Decision Analysis (MCDA), and Cluster Analysis (CA). Their findings revealed a growing trend toward hybrid analytical frameworks capable of supporting efficiency assessment, ranking, and decision-making simultaneously. The authors noted that clustering methods could improve DEA performance by grouping similar decision-making units before efficiency evaluation, thereby reducing heterogeneity within datasets.

Another significant development involved the application of machine learning for handling large-scale and synthetic datasets. Lychev (2023) proposed a synthetic data generation approach specifically designed for DEA studies. The research addressed the challenge of limited real-world datasets by generating artificial efficient and inefficient units for large-scale DEA experimentation. This contribution was particularly relevant for ML-enhanced DEA research, where large datasets are often required for model training, validation, and benchmarking purposes.

A major breakthrough in DEA-ML integration was reported by Li and colleagues (2024), who proposed a hybrid framework combining Data Envelopment Analysis with Recurrent Neural Networks (RNNs), including Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) architectures. Their model extended traditional DEA by forecasting future input and output values and subsequently evaluating future efficiency using DEA-Malmquist indices. Applied to research institutions within the Chinese Academy of Sciences, the framework demonstrated superior predictive performance compared with conventional regression-based approaches. The study illustrated how deep learning methods can transform DEA from a purely retrospective evaluation tool into a predictive decision-support system.

In parallel, broader research on hybrid optimization and machine learning approaches further strengthened the theoretical foundation for DEA-ML integration. Azevedo, Rocha, and Pereira (2024) conducted a comprehensive systematic literature review of hybrid optimization and machine learning methods. Although not exclusively focused on DEA, their findings emphasized that combining optimization-based techniques with machine learning can overcome the individual limitations of each approach. Their conclusions support the growing adoption of hybrid DEA-ML frameworks capable of handling increasingly complex operational and managerial problems.

Recent studies have also explored the integration of causal inference and advanced analytical modeling within DEA frameworks. Fukuyama, Tsionas, and Tan (2024) proposed incorporating causal modeling into DEA for performance evaluation. Their work moved beyond traditional correlation-based efficiency assessment by considering causal relationships among variables. Although not strictly a machine learning application, the study reflects a broader trend toward integrating advanced analytical and AI-inspired methodologies with DEA to improve interpretability and decision relevance.

Although research on ML-enhanced DEA has grown rapidly in recent years, the existing literature remains fragmented across different machine learning algorithms, DEA variants, and application domains. Studies have employed various approaches such as Artificial Neural Networks, Support Vector Machines, Random Forests, Gradient Boosting methods, Deep Learning models, and clustering algorithms in conjunction with DEA(Zhu et al., 2021). However, there is still limited consensus regarding the most effective integration strategies, the comparative advantages of different machine learning techniques, and the challenges associated with implementing hybrid DEA-ML frameworks. Furthermore, emerging issues related to explainability, computational complexity, data quality, model interpretability, and scalability require further investigation.

Therefore, this study aims to provide a comprehensive review of current developments in the integration of Machine Learning techniques with Data Envelopment Analysis models. Specifically, the study seeks to identify the major machine learning approaches applied in DEA research, examine their benefits and limitations, explore application areas across various industries, and highlight emerging trends and future research opportunities. By synthesizing recent advancements in this field, this review contributes to a deeper understanding of how machine learning can enhance efficiency analysis and support the development of next-generation performance evaluation frameworks.

Based on these objectives, the study addresses the following research questions: (1) How has machine learning been integrated into DEA models? (2) What machine learning algorithms are most frequently combined with DEA? (3) What advantages and limitations arise from the integration of machine learning and DEA? and (4) What future challenges and research opportunities remain for the development of hybrid DEA-ML approaches? Answering these questions will provide valuable insights for researchers, practitioners, and policymakers seeking to leverage both efficiency analysis and artificial intelligence technologies in complex decision-making environments.

2. Research Methodolgy

This study employs a Systematic Literature Review (SLR) approach to examine the current developments and future challenges associated with the integration of Machine Learning (ML) techniques into Data Envelopment Analysis (DEA) models(Charles et al., 2021). The systematic review methodology was selected because it provides a transparent, rigorous, and reproducible process for identifying, evaluating, and synthesizing relevant literature. Unlike traditional narrative reviews, an SLR minimizes researcher bias through predefined search procedures and selection criteria, thereby enabling a comprehensive understanding of the evolution of DEA-ML integration. To complement the systematic review, elements of bibliometric analysis and scoping review methodologies were incorporated to identify publication trends, emerging research themes, frequently used machine learning techniques, and future research directions. The combination of these approaches allows for both quantitative and qualitative examination of the literature while providing a broad overview of the field.

The literature search was conducted across several major academic databases recognized for their extensive coverage of scientific publications in operations research, artificial intelligence, management science, and engineering. These databases included Scopus, Web of Science, ScienceDirect, IEEE Xplore, and SpringerLink. The selection of multiple databases was intended to maximize literature coverage and reduce the possibility of omitting relevant studies. A structured search strategy was developed using combinations of keywords related to Data Envelopment Analysis and Machine Learning. The primary search terms included “Data Envelopment Analysis,” “Machine Learning,” “Artificial Intelligence,” “DEA and Neural Network,” “DEA and Deep Learning,” “DEA Optimization,” “Hybrid DEA Machine Learning,” “DEA Prediction Model,” “Efficiency Analysis and Artificial Intelligence,” and “Machine Learning-Based Efficiency Evaluation.” Boolean operators such as AND, OR, and NOT were employed to refine search results and improve retrieval accuracy. Additional relevant studies were identified through backward and forward citation tracking of key articles.

To ensure the quality and relevance of the selected literature, a set of inclusion and exclusion criteria was established before the screening process. Studies were included if they met the following conditions: (1) published between 2015 and 2026; (2) appeared in peer-reviewed journals, conference proceedings, or recognized academic publications; (3) written in English; (4) explicitly discussed the integration, application, or comparison of Machine Learning techniques with DEA models; and (5) provided sufficient methodological and empirical details for analysis. Studies were excluded if they were unpublished manuscripts, editorials, book reviews, duplicate records, non-English publications, or research focusing solely on DEA or Machine Learning without any meaningful integration between the two approaches. Articles with insufficient methodological information or limited relevance to efficiency analysis were also excluded from the final review.

The study selection process consisted of several stages. First, all records retrieved from the selected databases were collected and duplicate entries were removed. Second, titles and abstracts were screened to assess their relevance to the research objectives(O'Mara-Eves et al., 2015). Third, the full texts of potentially eligible studies were reviewed in detail to determine whether they satisfied the predefined inclusion criteria. Finally, the selected studies were compiled into a final dataset for detailed analysis and synthesis. This multi-stage screening process ensured that only high-quality and relevant studies were included in the review.

Following article selection, a structured data extraction procedure was performed to collect key information from each study. The extracted data included author names, publication year, country or region of application, application sector, DEA model employed, machine learning algorithm utilized, integration methodology, dataset characteristics, primary findings, reported advantages, identified limitations, and suggested future research directions. Collecting these variables enabled systematic comparison across studies and facilitated identification of recurring patterns and emerging trends within the literature. Furthermore, data extraction provided the basis for categorizing studies according to methodological and application-oriented characteristics.

To synthesize the findings, an analytical framework was developed to classify the selected studies into several categories(Onwuegbuzie & Weinbaum, 2017). The first classification focused on the type of Machine Learning technique employed, including Artificial Neural Networks (ANN), Support Vector Machines (SVM), Random Forests (RF), Gradient Boosting algorithms, Deep Learning models, Clustering techniques, Reinforcement Learning approaches, and other artificial intelligence methods. The second classification examined the DEA model used in each study, such as the CCR model, BCC model, Network DEA, Dynamic DEA, Stochastic DEA, DEA-Malmquist Index, and hybrid DEA variants. The third classification considered the application domain, including healthcare, banking and finance, manufacturing, supply chain management, transportation, energy systems, agriculture, education, and public sector performance evaluation. The fourth classification focused on the purpose of integration, such as feature selection, efficiency prediction, clustering and segmentation, outlier detection, performance forecasting, decision support, optimization, and efficiency interpretation.

The extracted data were subsequently analyzed using both descriptive and thematic approaches(Yukhymenko et al., 2014). Descriptive analysis was used to identify publication trends, commonly used machine learning algorithms, frequently applied DEA models, and dominant application sectors. Thematic analysis was employed to explore the benefits, limitations, methodological challenges, and future opportunities associated with DEA-ML integration. Particular attention was given to emerging research themes such as explainable artificial intelligence, deep learning-based efficiency analysis, federated learning, real-time efficiency monitoring, big data analytics, and intelligent decision support systems. Through this analytical process, the study aimed to provide a comprehensive understanding of how Machine Learning has enhanced traditional DEA methodologies and to identify promising directions for future research.

3. Results and Discussion

3.1 Publication Trends

The systematic review revealed a substantial increase in research interest regarding the integration of Machine Learning (ML) techniques with Data Envelopment Analysis (DEA) models over the past decade. Although studies combining DEA and artificial intelligence techniques existed prior to 2015, the volume of publications remained relatively limited and primarily focused on basic applications of Artificial Neural Networks (ANN) and Support Vector Machines (SVM) for efficiency prediction and classification. Between 2015 and 2018, publication activity increased gradually as researchers began recognizing the potential of machine learning methods to address several limitations of traditional DEA models, including variable selection, sensitivity to outliers, and predictive capability.

A more significant growth trend emerged after 2019, coinciding with the rapid advancement of artificial intelligence technologies and the widespread availability of large-scale datasets across various industries(Jacobides et al., 2021). During this period, researchers increasingly explored hybrid DEA-ML frameworks that combined efficiency measurement with predictive analytics, clustering techniques, and deep learning models. The growing popularity of Industry 4.0 technologies, big data analytics, and intelligent decision support systems further accelerated research activities in this area. As organizations generated increasingly complex and high-dimensional datasets, the need for advanced analytical methods capable of extracting meaningful insights from efficiency assessments became more apparent.

The review indicates that publication growth became particularly pronounced between 2021 and 2025. During these years, numerous studies introduced innovative frameworks that integrated DEA with advanced machine learning techniques such as Random Forests, Gradient Boosting algorithms, Deep Neural Networks, Long Short-Term Memory (LSTM) networks, and ensemble learning approaches. Researchers expanded the application of these hybrid methods beyond traditional sectors such as banking and manufacturing into healthcare, energy management, transportation systems, supply chain optimization, higher education, and environmental sustainability assessment. This diversification of application domains contributed significantly to the rapid expansion of the DEA-ML literature.

The increasing number of publications can also be attributed to the growing recognition that traditional DEA models alone may be insufficient for addressing contemporary analytical challenges(Emrouznejad & Yang, 2018). While DEA remains a powerful tool for measuring relative efficiency, it often struggles to handle nonlinear relationships, large-scale datasets, uncertainty, and dynamic operational environments. Machine learning techniques provide complementary capabilities that enhance the flexibility, scalability, and predictive power of DEA-based analyses. Consequently, researchers have increasingly adopted hybrid approaches that leverage the strengths of both methodologies.

An examination of geographical research distribution reveals that several countries have emerged as major contributors to DEA-ML integration research. China appears to be the leading contributor, accounting for a substantial proportion of recent publications. The country's strong investment in artificial intelligence research, digital transformation initiatives, and operational optimization has fostered extensive applications of DEA-ML models in manufacturing, energy systems, transportation networks, healthcare institutions, and research performance evaluation. Chinese universities and research institutes have produced numerous studies focusing on deep learning-enhanced DEA models, predictive efficiency analysis, and intelligent resource allocation systems.

The United States represents another major contributor to this research field. American researchers have focused primarily on methodological advancements, machine learning applications, healthcare efficiency evaluation, and business analytics. Strong interdisciplinary collaboration among departments of operations research, computer science, industrial engineering, and business administration has facilitated the development of innovative DEA-ML frameworks. Furthermore, the availability of large public datasets and advanced computational resources has supported extensive experimentation with machine learning-based efficiency analysis.

European countries have also played a significant role in advancing DEA-ML research(Göckenjan et al., 2009). The United Kingdom, Spain, Italy, Germany, and the Netherlands have produced influential studies on efficiency measurement, hybrid optimization techniques, and explainable artificial intelligence applications. European researchers have particularly emphasized methodological rigor, sustainability assessment, public sector performance evaluation, and healthcare efficiency analysis. Collaborative research projects funded by regional and international organizations have further contributed to the growth of this field across Europe.

In Asia, countries such as South Korea, Japan, India, and Taiwan have demonstrated increasing research activity related to DEA and machine learning integration. Researchers in these countries have explored applications in manufacturing systems, smart factories, energy management, supply chain optimization, and technology innovation assessment. The growing adoption of Industry 4.0 technologies and digital transformation strategies throughout Asia has created favorable conditions for the development of advanced efficiency evaluation frameworks.

Analysis of institutional contributions indicates that leading universities and research centers specializing in operations research, management science, artificial intelligence, and industrial engineering have been particularly active in this domain. Institutions with strong interdisciplinary research capabilities have demonstrated a greater tendency to develop hybrid DEA-ML frameworks.

Universities in China, the United States, and Europe consistently appear among the most productive institutions due to their substantial research funding, international collaborations, and expertise in both efficiency analysis and machine learning methodologies.

3.2 Types of Machine Learning Used with DEA

The integration of Machine Learning (ML) techniques with Data Envelopment Analysis (DEA) has evolved considerably over the past decade, resulting in the development of various hybrid frameworks designed to enhance efficiency evaluation and decision-making processes. Different machine learning algorithms contribute distinct capabilities to DEA applications, including prediction, classification, feature selection, clustering, and pattern recognition. The literature reveals that several categories of machine learning techniques have become particularly prominent in DEA-related research, including Artificial Neural Networks (ANN), Support Vector Machines (SVM), Random Forests (RF), Gradient Boosting methods, Deep Learning models, and Clustering Algorithms.

Among the earliest and most frequently utilized machine learning techniques in DEA integration are Artificial Neural Networks (ANNs). Inspired by the structure and functioning of biological neural systems, ANNs are capable of modeling complex nonlinear relationships between inputs and outputs. In DEA applications, neural networks are commonly employed for efficiency prediction and DEA score estimation. Researchers frequently use DEA efficiency scores as target variables for neural network models, enabling the prediction of future efficiency levels under varying operational conditions. This approach allows organizations to move beyond static efficiency assessment toward predictive performance evaluation. Furthermore, ANNs can capture nonlinear relationships that traditional DEA models may overlook, thereby improving the accuracy and robustness of efficiency analysis. Their adaptability and predictive capability have led to widespread applications in manufacturing, banking, healthcare, transportation, and energy management sectors.

Another widely adopted technique is the Support Vector Machine (SVM), which has demonstrated strong performance in classification and prediction tasks. In DEA-related studies, SVMs are commonly used to classify Decision-Making Units (DMUs) into efficient and inefficient categories based on efficiency scores generated by DEA models(Barr, 2004). By constructing optimal decision boundaries within high-dimensional spaces, SVMs can effectively distinguish between different efficiency groups even when datasets contain complex relationships. Researchers have also utilized SVMs to predict efficiency rankings and identify critical factors influencing organizational performance. Compared with traditional statistical classification methods, SVMs often provide higher classification accuracy and better generalization capabilities, making them particularly suitable for efficiency benchmarking and decision-support applications.

Random Forest (RF) algorithms have also gained considerable attention in DEA-machine learning integration due to their ability to handle large datasets and evaluate variable importance. Random Forest is an ensemble learning method that combines multiple decision trees to improve prediction accuracy and reduce overfitting(Parmar et al., 2018). Within DEA studies, Random Forests are frequently employed for feature selection and variable importance analysis. One of the key challenges in DEA is determining the most relevant input and output variables, as inappropriate variable selection can significantly affect efficiency scores. Random Forest algorithms address this challenge by identifying variables that contribute most significantly to efficiency outcomes. This capability enables researchers to simplify DEA models, reduce dimensionality, and improve interpretability while maintaining analytical accuracy. Additionally, Random Forests have been used to predict DEA efficiency scores and assess the relative impact of operational factors on organizational performance.

In recent years, Gradient Boosting algorithms, particularly Extreme Gradient Boosting (XGBoost), have emerged as powerful tools for predictive DEA applications. Gradient Boosting methods improve prediction performance by sequentially combining weak learners and correcting errors from previous iterations. XGBoost, in particular, has become popular due to its computational efficiency, scalability, and high predictive accuracy(Chen & Guestrin, 2016). In DEA-related research, XGBoost models are often trained using DEA efficiency scores as target variables to develop predictive efficiency assessment systems. These models can estimate future efficiency levels, identify influential performance drivers, and support strategic planning initiatives. Compared with traditional regression approaches, Gradient Boosting algorithms frequently achieve superior predictive performance, especially when dealing with complex and nonlinear datasets. Consequently, they have become increasingly prevalent in sectors such as finance, healthcare, manufacturing, and supply chain management.

The advancement of artificial intelligence technologies has further encouraged the integration of Deep Learning techniques with DEA frameworks. Deep Learning extends traditional neural networks by incorporating multiple hidden layers capable of learning hierarchical representations from large and complex datasets(Deng, 2014). Deep Learning models, including Deep Neural Networks (DNNs), Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and Gated Recurrent Units (GRUs), have demonstrated remarkable success in pattern recognition, forecasting, and large-scale data analysis. In DEA applications, Deep Learning is particularly valuable for handling high-dimensional data, temporal efficiency analysis, and complex nonlinear relationships. Researchers have applied deep learning models to predict future efficiency scores, forecast organizational performance, and analyze dynamic production systems. Furthermore, recurrent architectures such as LSTM and GRU have enabled the development of predictive DEA frameworks capable of incorporating time-series data and evaluating future efficiency trends. As organizations continue to generate increasingly large and diverse datasets, Deep Learning is expected to play an increasingly important role in next-generation DEA methodologies.

In addition to predictive and classification algorithms, Clustering Algorithms have become important tools for enhancing DEA analyses. Clustering techniques are unsupervised machine learning methods that group similar observations based on their characteristics(Kassambara, 2017). In DEA research, clustering algorithms are commonly applied before efficiency evaluation to create homogeneous groups of DMUs. This preprocessing step addresses one of the key limitations of DEA, namely the assumption that all decision-making units are directly comparable. By grouping similar entities prior to analysis, researchers can reduce heterogeneity and improve the reliability of efficiency comparisons. Among the most widely used clustering methods are K-Means Clustering and Hierarchical Clustering. K-Means partitions DMUs into predefined clusters based on similarity measures, while Hierarchical Clustering constructs nested groupings without requiring a predetermined number of clusters. These techniques have been successfully applied in healthcare, banking, education, energy management, and manufacturing sectors to improve benchmarking accuracy and facilitate more meaningful efficiency evaluations.

3.3 Integration Frameworks

The growing convergence of Machine Learning (ML) and Data Envelopment Analysis (DEA) has led to the development of several integration frameworks designed to enhance the effectiveness, robustness, and predictive capabilities of efficiency analysis. While traditional DEA provides a powerful mechanism for measuring the relative efficiency of decision-making units (DMUs), it is often limited by challenges related to variable selection, data quality, outlier sensitivity, and predictive functionality(de Sousa et al., 2005). Machine learning techniques address many of these limitations by offering advanced capabilities for data preprocessing, pattern recognition, forecasting, and decision support.

The Pre-DEA Machine Learning framework involves applying machine learning techniques before conducting DEA analysis. In this approach, machine learning serves as a preprocessing tool that improves the quality and suitability of data used in efficiency measurement. Since DEA outcomes are highly dependent on the selection of input and output variables, data quality, and the presence of outliers, preprocessing plays a crucial role in ensuring reliable efficiency estimates.

One of the most common applications of Pre-DEA ML is feature selection. Traditional DEA models often rely on expert judgment to determine appropriate input and output variables, which can introduce subjectivity and potentially affect the accuracy of efficiency scores. Machine learning algorithms such as Random Forest, LASSO Regression, Recursive Feature Elimination, and Gradient Boosting can identify the most relevant variables by evaluating their contribution to performance outcomes. By eliminating redundant or irrelevant variables, these techniques improve model simplicity, reduce dimensionality, and enhance DEA robustness.

Another important application involves data preprocessing and dimensionality reduction. Large datasets frequently contain missing values, noisy observations, and highly correlated variables that may distort DEA results. Machine learning techniques such as Principal Component Analysis (PCA), Autoencoders, and Data Normalization methods can transform raw data into more structured and meaningful representations before efficiency evaluation(Alkhayrat et al., 2020). These preprocessing methods help reduce noise, improve comparability among DMUs, and facilitate more reliable efficiency measurement.

Outlier detection is also a significant component of the Pre-DEA framework. DEA is particularly sensitive to extreme observations because efficient frontier construction depends directly on observed data points. Outliers can artificially influence the efficiency frontier and lead to misleading efficiency scores. Machine learning algorithms such as Isolation Forest, Local Outlier Factor (LOF), and clustering-based anomaly detection methods can identify and remove abnormal observations before DEA implementation. Consequently, the resulting efficiency estimates become more stable and representative of actual organizational performance.

The primary advantage of the Pre-DEA ML framework lies in its ability to improve data quality and model validity prior to efficiency analysis(Nguyen, 2015). By addressing data-related challenges at the initial stage, researchers can obtain more accurate and reliable DEA outcomes while reducing the impact of methodological biases.

The DEA-ML Hybrid Framework represents the most comprehensive form of integration between machine learning and DEA. In this approach, DEA and ML techniques are combined within a unified analytical process, where DEA outputs frequently serve as inputs for machine learning models or vice versa. Rather than functioning solely as a preprocessing or post-analysis tool, machine learning becomes an integral component of the efficiency evaluation framework.

The literature indicates that many contemporary studies employ a multi-stage integration process that combines elements of all three frameworks(Wiese et al., 2020). A typical workflow may involve machine learning-based feature selection and preprocessing, followed by DEA efficiency analysis, and finally machine learning-based prediction or interpretation of efficiency outcomes. This process can be represented as follows:

Inputs and Outputs → Machine Learning Feature Selection and Data Preprocessing → DEA Efficiency Evaluation → Efficiency Scores → Machine Learning Prediction and Interpretation

This integrated workflow demonstrates how machine learning can support every stage of the DEA process, from data preparation to efficiency evaluation and strategic decision-making. As machine learning technologies continue to evolve, future integration frameworks are expected to become increasingly intelligent, adaptive, and explainable, enabling more sophisticated forms of efficiency analysis across diverse organizational and industrial contexts.

3.4 Application Areas of DEA–Machine Learning Integration

The integration of Data Envelopment Analysis (DEA) and Machine Learning (ML) has expanded significantly across a wide range of application domains. The complementary strengths of DEA in measuring relative efficiency and ML in prediction, classification, and pattern recognition have enabled organizations to address increasingly complex performance evaluation challenges. The literature indicates that DEA-ML frameworks have been successfully applied in healthcare, banking and finance, supply chain management, manufacturing, energy systems, agriculture, and higher education. These applications demonstrate the versatility and growing importance of hybrid analytical approaches in supporting data-driven decision-making and operational improvement.

The healthcare sector represents one of the most prominent application areas for DEA-ML integration(Brazier, 2008). Hospitals and healthcare organizations operate in highly complex environments characterized by multiple inputs, diverse service outputs, limited resources, and increasing demand for quality care. Traditional DEA models have long been used to evaluate hospital efficiency by comparing healthcare facilities based on factors such as medical staff, hospital beds, operational expenditures, patient admissions, and treatment outcomes. However, the increasing availability of electronic health records and healthcare big data has created opportunities for incorporating machine learning techniques into efficiency assessment frameworks.

In healthcare applications, machine learning algorithms are frequently employed to predict hospital efficiency scores, identify performance determinants, and classify hospitals according to operational effectiveness. Neural networks, Random Forests, and Support Vector Machines have been used to analyze relationships between hospital resources and healthcare outcomes. Additionally, clustering algorithms help group hospitals with similar characteristics before DEA evaluation, improving benchmarking accuracy. Recent studies have also utilized deep learning models to forecast future healthcare performance and resource utilization patterns. These hybrid approaches provide healthcare administrators with more comprehensive insights into operational efficiency and support evidence-based policy decisions.

Banking and financial institutions constitute another major application domain for DEA-ML integration(Jomthanachai et al., 2021). Financial organizations continuously seek to improve operational efficiency while managing risks, maximizing profitability, and maintaining regulatory compliance. DEA has traditionally been used to evaluate bank performance by analyzing the relationship between inputs such as labor, capital, operating expenses, and deposits and outputs such as loans, investments, and financial returns.

Machine learning techniques enhance these analyses by introducing predictive and risk assessment capabilities(Motwani et al., 2017). For example, DEA-generated efficiency scores can be used as target variables for machine learning models that predict future bank performance under different economic conditions. Random Forests, Support Vector Machines, and Gradient Boosting algorithms have been employed to classify banks according to efficiency levels and identify critical performance drivers. Furthermore, DEA-ML frameworks have demonstrated significant potential in credit risk analysis, where machine learning algorithms evaluate borrower behavior, financial indicators, and historical performance data to predict default risks. The combination of efficiency measurement and predictive analytics enables financial institutions to optimize resource allocation, improve lending decisions, and strengthen risk management practices.

The increasing complexity of global supply chains has created substantial demand for advanced performance evaluation methods(Gunasekaran & Kobu, 2007). DEA-ML integration has emerged as a valuable tool for assessing logistics efficiency, transportation performance, warehouse operations, and overall supply chain effectiveness. DEA models are commonly used to evaluate the efficiency of supply chain entities based on inputs such as transportation costs, labor resources, inventory levels, and operational expenditures, while outputs may include delivery performance, customer satisfaction, and service quality indicators.

Machine learning techniques complement DEA by enabling predictive logistics management and operational optimization. Clustering algorithms can segment supply chain entities according to operational characteristics, allowing more meaningful efficiency comparisons. Random Forests and Gradient Boosting methods assist in identifying key performance drivers and predicting logistics outcomes. Deep learning models have also been applied to forecast transportation demand, inventory requirements, and delivery performance. These capabilities enable organizations to proactively manage supply chain disruptions, improve operational resilience, and enhance overall efficiency.

Manufacturing remains one of the most active sectors for DEA-ML applications, particularly within the context of Industry 4.0 and smart factory initiatives. Modern manufacturing systems generate vast amounts of data from sensors, production equipment, enterprise systems, and Internet of Things (IoT) devices. While DEA provides a structured framework for evaluating production efficiency, machine learning enables organizations to exploit these large datasets for predictive and adaptive decision-making.

In manufacturing environments, DEA-ML models are commonly used to assess production line efficiency, equipment utilization, product quality, and resource allocation effectiveness. Machine learning algorithms such as Artificial Neural Networks, Random Forests, and Deep Learning models can predict future production performance and detect inefficiencies before they affect operational outcomes. Predictive maintenance applications represent another important area, where machine learning techniques analyze sensor data to anticipate equipment failures and optimize maintenance schedules. By integrating DEA efficiency assessment with machine learning-driven forecasting, smart factories can achieve higher productivity, reduced operational costs, and improved resource utilization.

The energy sector has increasingly adopted DEA-ML frameworks to address challenges related to sustainability, renewable energy development, and resource optimization. DEA has been widely used to evaluate the efficiency of power plants, energy distribution systems, renewable energy facilities, and national energy infrastructures. Traditional efficiency indicators often focus on inputs such as fuel consumption, capital investment, and labor resources, while outputs include electricity generation, energy distribution, and environmental performance measures.

Machine learning enhances energy efficiency analysis by supporting prediction, optimization, and scenario evaluation(Mosavi et al., 2019). Neural networks and deep learning models have been employed to forecast renewable energy generation from sources such as solar and wind power. Random Forests and Gradient Boosting algorithms help identify factors affecting energy efficiency and environmental sustainability. In renewable energy systems, DEA-ML models can evaluate operational efficiency while simultaneously predicting future energy production under varying weather conditions. These integrated frameworks support strategic planning, investment decisions, and sustainable energy management initiatives.

Agriculture represents another important area where DEA and machine learning techniques have been combined to improve productivity assessment and resource management. Agricultural production depends on numerous inputs, including land, labor, machinery, water, fertilizers, and technological resources, making efficiency evaluation particularly complex. DEA has traditionally been used to assess farm efficiency by comparing agricultural producers based on resource utilization and crop yields.

The incorporation of machine learning techniques enables more sophisticated analysis of agricultural performance. Machine learning algorithms can process large datasets generated from precision agriculture technologies, satellite imagery, sensor networks, and climate monitoring systems. These models assist in predicting crop yields, identifying productivity determinants, and optimizing resource allocation. Clustering techniques can group farms with similar production characteristics before DEA implementation, improving comparability and benchmarking accuracy. Furthermore, deep learning methods have demonstrated significant potential for integrating environmental and operational variables into efficiency assessments, thereby supporting sustainable agricultural development.

Higher education institutions increasingly face pressure to demonstrate efficiency, accountability, and academic performance. Consequently, DEA has become a widely used tool for evaluating university efficiency by examining the relationship between inputs such as faculty members, funding, infrastructure, and administrative resources and outputs such as research publications, graduation rates, student enrollment, and academic achievements.

Machine learning techniques extend these analyses by introducing predictive capabilities and advanced performance evaluation mechanisms. Researchers have applied Artificial Neural Networks, Support Vector Machines, and ensemble learning methods to predict institutional efficiency, identify factors influencing academic performance, and forecast research productivity. Clustering algorithms are frequently used to group universities with similar characteristics before DEA analysis, ensuring more meaningful efficiency comparisons. Additionally, machine learning models can analyze large educational datasets to uncover hidden patterns related to student outcomes, faculty performance, and resource utilization. These insights support strategic planning, institutional benchmarking, and evidence-based decision-making in higher education management.

3.5 Benefits of Machine Learning and Data Envelopment Analysis Integration

One of the most significant advantages of ML-DEA integration is improved feature selection. Traditional DEA models require researchers or decision-makers to determine appropriate input and output variables before conducting efficiency analysis. The inclusion of irrelevant, redundant, or highly correlated variables can negatively affect efficiency scores and reduce model reliability. Machine learning algorithms such as Random Forests, Recursive Feature Elimination, LASSO Regression, and Gradient Boosting techniques provide systematic approaches for identifying the most relevant variables within large datasets. By automatically selecting influential variables and removing unnecessary attributes, machine learning enhances the quality of DEA models and reduces the risk of biased efficiency evaluations. This capability is particularly valuable in complex environments where organizations collect large amounts of operational data from multiple sources.

Another major benefit of ML-DEA integration is higher prediction accuracy. Traditional DEA models are primarily designed for retrospective efficiency assessment and do not inherently provide predictive capabilities. Machine learning techniques address this limitation by enabling efficiency forecasting and future performance prediction. Artificial Neural Networks, Support Vector Machines, Random Forests, and Deep Learning models can be trained using historical DEA efficiency scores to predict future organizational performance. This predictive capability allows decision-makers to anticipate efficiency changes, evaluate alternative scenarios, and implement proactive improvement strategies. Consequently, DEA evolves from a static performance measurement tool into a dynamic decision-support system capable of supporting strategic planning and resource optimization.

The ability to handle large and complex datasets represents another important advantage of integrating machine learning with DEA. In modern organizational environments, vast amounts of data are generated through digital technologies, enterprise systems, sensors, and Internet of Things (IoT) devices. Traditional DEA models may encounter difficulties when analyzing high-dimensional datasets characterized by numerous variables and observations. Machine learning algorithms are specifically designed to process large-scale data efficiently and can identify meaningful patterns within complex datasets. Through dimensionality reduction, feature extraction, and automated data processing techniques, ML enhances the scalability of DEA applications and enables efficiency analysis in big data environments. This capability has become increasingly important in sectors such as healthcare, manufacturing, finance, and energy management, where data volumes continue to grow rapidly.

A further advantage of ML-DEA integration is its ability to model nonlinear relationships. Conventional DEA assumes linear production frontiers and may struggle to capture complex interactions among inputs and outputs. However, real-world operational systems often exhibit nonlinear behavior due to technological, environmental, economic, and managerial factors. Machine learning algorithms, particularly Artificial Neural Networks and Deep Learning models, excel at identifying and modeling nonlinear relationships within data. By incorporating these techniques into DEA frameworks, researchers can develop more realistic and accurate representations of organizational performance. The ability to capture complex patterns improves the explanatory power of efficiency analyses and provides deeper insights into the factors influencing operational outcomes.

Automation constitutes another significant benefit of integrating machine learning and DEA. Traditional DEA implementation often requires extensive manual intervention, including data preparation, variable selection, outlier detection, model specification, and result interpretation. These processes can be time-consuming and subject to human bias. Machine learning techniques automate many of these tasks by utilizing intelligent algorithms capable of preprocessing data, detecting anomalies, selecting variables, and generating predictive models with minimal human involvement. Automation not only increases efficiency and consistency but also enables organizations to conduct real-time or near-real-time performance assessments. As a result, managers can receive timely information and make faster decisions in rapidly changing operational environments.

In addition to these core benefits, ML-DEA integration improves robustness against data uncertainty and noise. Traditional DEA models are sensitive to extreme observations and measurement errors because efficiency frontiers are constructed directly from observed data. Machine learning techniques provide advanced mechanisms for outlier detection, anomaly identification, and noise reduction, thereby improving the stability and reliability of efficiency estimates. Furthermore, machine learning algorithms can accommodate incomplete or partially missing datasets through imputation and data recovery techniques, enhancing the practicality of DEA applications in real-world settings.

The integration also facilitates enhanced decision support capabilities. DEA identifies efficient and inefficient units, but it does not always explain why efficiency differences occur or how performance can be improved. Machine learning techniques contribute additional analytical depth by uncovering hidden patterns, identifying key performance drivers, and generating predictive insights. Decision-makers can therefore move beyond simple efficiency rankings and gain a more comprehensive understanding of organizational behavior. Such information is valuable for strategic planning, resource allocation, performance benchmarking, and continuous improvement initiatives.

Another emerging benefit involves the development of intelligent and adaptive efficiency evaluation systems. As machine learning models continuously learn from new data, DEA-ML frameworks can adapt to changing operational conditions and evolving business environments. This adaptability is particularly important in industries characterized by rapid technological advancement, fluctuating market conditions, and dynamic resource requirements. Adaptive DEA-ML systems provide organizations with more flexible and responsive performance management tools compared with traditional static efficiency evaluation approaches.

3.6 Current Challenges of Machine Learning and Data Envelopment Analysis Integration

One of the most significant challenges in DEA-ML integration is the quality of available data. Both DEA and machine learning algorithms rely heavily on accurate, complete, and representative datasets. However, real-world datasets frequently contain missing values, inconsistent records, measurement errors, and noisy observations that can negatively affect analytical outcomes.

Missing values represent a common problem across many application domains, including healthcare, finance, manufacturing, and education. Incomplete information may distort efficiency measurements and reduce the predictive accuracy of machine learning models. Although various imputation techniques can be employed to estimate missing data, inappropriate imputation strategies may introduce bias and compromise model reliability. Consequently, researchers must carefully select suitable data preprocessing techniques to ensure the validity of DEA-ML analyses.

Another important issue relates to imbalanced datasets. In many efficiency evaluation scenarios, the number of efficient decision-making units (DMUs) is significantly smaller than the number of inefficient units. Such imbalances can affect the performance of classification and prediction algorithms, leading to biased model outcomes and reduced predictive accuracy. Machine learning models trained on imbalanced datasets may struggle to accurately identify efficient units, limiting their usefulness for decision-support purposes. Addressing this challenge often requires advanced sampling techniques, data augmentation methods, or specialized machine learning algorithms capable of handling class imbalance.

Furthermore, data heterogeneity remains a persistent concern. Organizations operating in different environments may collect data using varying standards, measurement units, and reporting procedures. This inconsistency complicates data integration and reduces the comparability of efficiency analyses across different studies and industries.

Interpretability is another major limitation of contemporary DEA-ML frameworks. Traditional DEA models are generally considered transparent because efficiency scores are calculated through well-defined mathematical optimization procedures. Decision-makers can readily understand the relationships between inputs, outputs, and efficiency outcomes. In contrast, many machine learning algorithms, particularly deep learning models, operate as complex "black-box" systems whose internal decision-making processes are difficult to interpret.

Artificial Neural Networks, Deep Neural Networks, Long Short-Term Memory (LSTM) networks, and other advanced machine learning architectures often achieve high predictive accuracy but provide limited insight into how predictions are generated. This lack of transparency creates challenges for managers, policymakers, and stakeholders who require understandable explanations for efficiency evaluations and strategic decisions. In sectors such as healthcare, finance, and public administration, where accountability and regulatory compliance are critical, limited interpretability can significantly hinder the acceptance and implementation of DEA-ML solutions.

Moreover, difficulties in explaining model behavior reduce trust in machine learning-based efficiency assessments. Organizations may hesitate to adopt complex predictive models if they cannot clearly understand the factors influencing efficiency scores and recommendations.

The increasing sophistication of machine learning algorithms has also introduced substantial computational challenges. Traditional DEA models already require solving multiple optimization problems, particularly when evaluating large numbers of decision-making units. When combined with computationally intensive machine learning techniques, the overall analytical process can become resource-demanding and time-consuming.

Large-scale datasets frequently contain thousands of observations and numerous variables, requiring significant processing power and memory resources. Deep learning models, ensemble learning methods, and advanced optimization algorithms often involve extensive training procedures that demand specialized hardware such as Graphics Processing Units (GPUs) or high-performance computing environments. Consequently, organizations with limited computational resources may face difficulties implementing advanced DEA-ML frameworks.

Computational complexity becomes even more pronounced in real-time applications where efficiency assessments must be continuously updated. Dynamic production systems, smart factories, financial markets, and energy management platforms require rapid analysis and decision-making capabilities. Developing DEA-ML models that can efficiently process large-scale data while maintaining high predictive accuracy remains a significant research challenge.

A further limitation concerns the generalizability of DEA-ML models across different industries and operational contexts. Many existing studies focus on specific sectors such as healthcare, banking, manufacturing, or education. While models often perform well within the environments for which they were developed, their effectiveness may decline when applied to different industries or geographical regions.

This challenge arises because efficiency determinants vary substantially across sectors. Variables that strongly influence efficiency in manufacturing may have limited relevance in healthcare or financial services. Similarly, organizational structures, regulatory environments, and operational processes differ significantly across industries. As a result, machine learning models trained on one dataset may not accurately predict efficiency outcomes in another context.

The issue of generalizability also affects international comparisons. Differences in economic conditions, technological development, institutional frameworks, and cultural factors can influence organizational performance and efficiency measurement. Consequently, DEA-ML models often require substantial adaptation before being applied to new environments, limiting their transferability and scalability.

The growing complexity of machine learning models has intensified the need for Explainable Artificial Intelligence (XAI) within DEA-ML research. Explainability refers to the ability of a model to provide understandable and transparent explanations for its predictions and decisions. While predictive accuracy remains important, stakeholders increasingly demand models that can clearly justify efficiency assessments and performance recommendations.

Current DEA-ML frameworks often prioritize predictive performance over interpretability, creating a trade-off between accuracy and transparency. This challenge is particularly important in high-stakes decision-making environments where efficiency evaluations influence resource allocation, investment decisions, policy development, and organizational strategy.

Emerging XAI techniques such as SHAP (Shapley Additive Explanations), LIME (Local Interpretable Model-Agnostic Explanations), feature importance analysis, and attention mechanisms offer promising solutions for improving transparency. However, the integration of these methods into DEA-ML frameworks remains relatively limited. Future research must focus on developing explainable hybrid models that balance predictive power with interpretability, thereby enhancing stakeholder confidence and facilitating practical implementation.

One of the most fundamental challenges in DEA-ML integration is the absence of standardized methodological frameworks. Existing studies employ a wide variety of DEA models, machine learning algorithms, data preprocessing techniques, evaluation metrics, and integration strategies. While this diversity reflects the flexibility of hybrid approaches, it also creates difficulties in comparing results across studies and identifying best practices.

Researchers have proposed numerous integration frameworks, including Pre-DEA Machine Learning, DEA-ML Hybrid Models, and Post-DEA Machine Learning approaches. However, there is currently no universally accepted methodology regarding how machine learning should be incorporated into DEA analyses. Different studies often use different datasets, variable selection procedures, validation methods, and performance metrics, making it difficult to evaluate the relative effectiveness of alternative approaches.

The lack of standardization also affects reproducibility. Without common methodological guidelines, researchers may encounter difficulties replicating previous studies or validating reported findings. Establishing standardized frameworks for data preparation, model development, evaluation, and reporting would significantly improve the consistency and credibility of DEA-ML research.

4. Conclusion

This study reviewed the current developments and future challenges associated with the integration of Machine Learning (ML) and Data Envelopment Analysis (DEA), highlighting the growing importance of hybrid analytical frameworks for efficiency evaluation and decision support. The findings indicate that ML integration significantly enhances traditional DEA capabilities by improving feature selection, predictive accuracy, scalability, automation, and the ability to model complex nonlinear relationships. Among the various machine learning techniques, Artificial Neural Networks (ANN), Support Vector Machines (SVM), Random Forests, Gradient Boosting methods, and Deep Learning models emerged as the most frequently adopted approaches for efficiency prediction, classification, feature selection, and performance analysis. The review also revealed that DEA-ML applications have expanded across diverse sectors, including healthcare, banking and finance, manufacturing, supply chain management, energy systems, agriculture, and higher education, demonstrating the versatility and practical value of these integrated methodologies. Furthermore, this study contributes to the existing literature by providing a comprehensive overview of DEA-ML integration frameworks, application areas, benefits, and current challenges while identifying major research trends and knowledge gaps. Despite the substantial progress achieved, issues related to data quality, model interpretability, computational complexity, generalizability, explainability, and methodological standardization continue to limit broader adoption. Looking forward, emerging research directions such as Explainable Artificial Intelligence (XAI), federated learning, deep learning-based efficiency analysis, real-time analytics, and intelligent decision-support systems offer significant opportunities for advancing the field. However, the development of transparent, robust, and standardized DEA-ML frameworks remains a critical challenge that must be addressed to ensure the reliability, scalability, and practical applicability of future efficiency evaluation systems.

References

Alkhayrat, M., Aljnidi, M., & Aljoumaa, K. (2020). A comparative dimensionality reduction study in telecom customer segmentation using deep learning and PCA. *Journal of Big Data*, 7(1), 9.

Barr, R. S. (2004). DEA software tools and technology: A state-of-the-art survey. In *Handbook on data envelopment analysis* (pp. 539–566). Springer.

Brazier, J. (2008). PMC15 in or Out? Empirical Evidence on Income Losses in Health State Valuations and Implications for Economic Evaluations. *Value in Health*.

Bughin, J., Hazan, E., Sree Ramaswamy, P., DC, W., & Chu, M. (2017). *Artificial intelligence the next digital frontier*.

Charles, V., Gherman, T., & Zhu, J. (2021). Data envelopment analysis and big data: A systematic literature review

with bibliometric analysis. *Data-Enabled Analytics*, 1–29.
Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 785–794.
Cooper, W. W., Seiford, L. M., Tone, K., & Zhu, J. (2007). Some models and measures for evaluating performances with DEA: past accomplishments and future prospects. *Journal of Productivity Analysis*, *28*(3), 151–163.
de Sousa, M. da C. S., Cribari-Neto, F., & Stosic, B. D. (2005). Explaining DEA technical efficiency scores in an outlier corrected environment: the case of public services in Brazilian municipalities. *Brazilian Review of Econometrics*, *25*(2), 287–313.
Deng, L. (2014). A tutorial survey of architectures, algorithms, and applications for deep learning. *APSIPA Transactions on Signal and Information Processing*, *3*, e2.
Emrouznejad, A., & Yang, G. (2018). A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016. *Socio-Economic Planning Sciences*, *61*, 4–8.
Göckenjan, G., Nunner-Winkler, G., Iser, M., Wittwer, H., Schnädelbach, H., Reemtsma, J. P., Soeffner, H., Spittler, G., Bruchhausen, W., & Baranzke, H. (2009). Begriffe: Homo faber/Technik. *Handbuch Anthropologie*.
Gunasekaran, A., & Kobu, B. (2007). Performance measures and metrics in logistics and supply chain management: a review of recent literature (1995–2004) for research and applications. *International Journal of Production Research*, *45*(12), 2819–2840.
Jacobides, M. G., Brusoni, S., & Candelon, F. (2021). The evolutionary dynamics of the artificial intelligence ecosystem. *Strategy Science*, *6*(4), 412–435.
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.
Kassambara, A. (2017). *Practical guide to cluster analysis in R: Unsupervised machine learning* (Vol. 1). Sthda.
Mosavi, A., Salimi, M., Faizollahzadeh Ardabili, S., Rabczuk, T., Shamshirband, S., & Varkonyi-Koczy, A. R. (2019). State of the art of machine learning models in energy systems, a systematic review. *Energies*, *12*(7), 1301.
Motwani, M., Dey, D., Berman, D. S., Germano, G., Achenbach, S., Al-Mallah, M. H., Andreini, D., Budoff, M. J., Cademartiri, F., & Callister, T. Q. (2017). Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. *European Heart Journal*, *38*(7), 500–507.
Nataraja, N. R., & Johnson, A. L. (2011). Guidelines for using variable selection techniques in data envelopment analysis. *European Journal of Operational Research*, *215*(3), 662–669.
Nguyen, N. M. (2015). *Restructuring an emerging market: The case of Vietnam*. Golden Gate University.
O'Mara-Eves, A., Thomas, J., McNaught, J., Miwa, M., & Ananiadou, S. (2015). Using text mining for study identification in systematic reviews: a systematic review of current approaches. *Systematic Reviews*, *4*(1), 5.
Onwuegbuzie, A. J., & Weinbaum, R. (2017). A framework for using qualitative comparative analysis for the review of the literature. *The Qualitative Report*, *22*(2), 359–372.
Parmar, A., Katariya, R., & Patel, V. (2018). A review on random forest: An ensemble classifier. *International Conference on Intelligent Data Communication Technologies and Internet of Things*, 758–763.
Wiese, L., Pohlmeyer, A. E., & Hekkert, P. (2020). Design for sustained wellbeing through positive activities—a multi-stage framework. *Multimodal Technologies and Interaction*, *4*(4), 71.
Yukhymenko, M. A., Brown, S. W., Lawless, K. A., Brodowinska, K., & Mullin, G. (2014). Thematic analysis of teacher instructional practices and student responses in middle school classrooms with problem-based learning environment. *Global Education Review*, *1*(3), 93–110.
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.