International Journal of Industrial Engineering & Production Research

Journal Website: http://ijieen.iust.ac.ir/

A Novel DEA-XGBoost Hybrid Framework for Superior Bank Branch Efficiency Assessment

Hamid Eslami Nosratabadi*, Orlando Belo

***Hamid Eslami Nosratabadi**, ALGORITMI/LASI Centre, School of Engineering, University of Minho, Braga, Portugal*

***Orlando Belo,** ALGORITMI/LASI Centre, School of Engineering, University of Minho, Braga, Portugal*

KEYWORDS

DEA;
XGBoost;
Bank Branch Performance;
Efficiency Evaluation;
Hybrid Machine Learning Model;
Decision Support System.

ABSTRACT

With the growing complexity of banking operations and increasing customer expectations, reliable evaluation of bank branch performance requires methods capable of integrating multidimensional indicators and identifying nonlinear performance patterns. This work proposes a hybrid DEA–machine-learning framework combining Data Envelopment Analysis DEA for relative efficiency assessment with supervised learning for performance classification. The framework is applied to real-world data from 3,124 bank branches of a major commercial bank in Iran, using 30 financial and non-financial key performance indicators. First, an input-oriented CCR DEA model is used to calculate branch-level efficiency scores. These continuous scores are then discretized into four performance classes Efficient, Relatively Efficient, Moderately Efficient, and Inefficient which constitute the target variable for subsequent classification. Several machine-learning models, including XGBoost, Support Vector Machine, Random Forest, Multilayer Perceptron, and a stacked ensemble, are trained and evaluated using an 80/20 train-test split and 10-fold cross-validation. Results show that SVM achieves the highest reported classification accuracy, while XGBoost provides competitive predictive performance together with interpretable feature-importance analysis. The proposed framework therefore integrates DEA-based benchmarking with machine-learning-based classification and diagnostic analysis, providing a practical decision-support approach for identifying performance differences and supporting targeted managerial interventions in bank branch networks.

1. Introduction

In recent years, the banking sector has undergone profound structural and technological transformations. On one hand, these changes are driven by digitalization and the proliferation of electronic service channels. On the other, they are intensified by competitive pressures, regulatory requirements, and heightened customer expectations. Despite the significant growth in digital transactions, the branch network remains a critical touchpoint for customers, particularly for delivering complex banking services. Consequently, evaluating branch performance continues to be a strategic priority for banks. Traditional efficiency assessment methods, such as Data Envelopment Analysis[1], have been

* Corresponding Author: Email: Hamideslami.na@gmail.com , Email: obelo@di.uminho.pt

widely adopted due to their ability to measure the relative efficiency of decision-making units[2] in the presence of multiple inputs and outputs. However, these methods face limitations in handling large datasets, nonlinear relationships, and non-financial or behavioral variables, particularly in terms of predictability and uncovering complex patterns [1]. In parallel, recent advancements in machine learning have enabled the use of techniques like XGBoost for modeling nonlinear relationships, scalability, and feature importance extraction. Yet, these methods alone lack the input-output-based benchmarking structure that DEA provides [2]. Therefore, this article addresses the intersection of DEA's capabilities in measuring relative efficiency with the predictive and explanatory power of advanced machine learning models (particularly XGBoost) to develop a hybrid framework that is both practical and reliable for evaluating and predicting bank branch efficiency [3].

Addressing the evaluation of bank branch efficiency through a DEA–XGBoost hybrid lens holds substantial theoretical and practical significance. From a managerial perspective, banks must make optimal decisions regarding resource allocation, branch network redesign, and human resource planning. These decisions are most effective when efficiency insights are provided not only descriptively but also predictively and interpretably to managers [4]. Furthermore, market regulators and supervisors require measurable indicators that enable tracking, comparison, and monitoring of efficiency changes at the branch level, making them receptive to reliable and transparent frameworks. On the other hand, the emergence of rich operational and customer-centric data (including digital metrics, customer satisfaction indices, and transactional exchanges) necessitates that efficiency evaluation models leverage these multidimensional datasets. In this context, integrating DEA with algorithms capable of modeling complexities and nonlinear interactions can bridge the gap between conventional efficiency measurement and advanced data-driven analysis [5]. Moreover, a decision support system based on such a framework can facilitate early identification of underperforming branches, prioritization of corrective interventions, and the design of incentive policies to enhance performance – a matter of growing importance in today's competitive banking landscape [6].

Despite notable progress in the application of DEA and machine learning algorithms in recent studies, three primary gaps persist in the literature, which this paper aims to address. First, most traditional DEA studies have focused on descriptive efficiency analysis and have not adequately covered predictive capabilities or the identification of causal factors leading to efficiency or inefficiency. In other words, translating efficiency scores into actionable knowledge for predictive and managerial responses remains unresolved [7]. Second, studies that integrate DEA with machine learning have often concentrated on low-dimensional, small-volume datasets or employed algorithms that offer limited scalability or interpretability. Thus, the integration of DEA's reference logic with modern boosting models like XGBoost, especially when dealing with large-scale operational network data, remains unclear [8]. Third, many research has centered on financial metrics, with less analysis of non-financial indicators (such as digital channel adoption rates, customer satisfaction, and operational efficiency of ATMs/POS devices). Notably, the interactions between these indicators and financial variables may exert hidden effects on efficiency that linear or semi-parametric models alone cannot reveal [9]. To fill these gaps, this work endeavors to propose a hybrid framework based on computing DEA scores and then training an XGBoost model (using selected financial and non-financial variables) to predict and explain efficiency classes, thereby emphasizing both benchmarking and predictive aspects [10].

The primary innovations of this research lie in three areas: first, the proposed hybrid architecture that leverages DEA's explanatory advantages as an input-driven benchmarking mechanism and XGBoost's nonlinear modeling and feature importance extraction to operationally explain the reasons and mechanisms of inefficiency; second, the simultaneous use of an extensive, multidimensional set of financial and non-financial branch metrics such as customer digitalization variables, satisfaction indices, and device operational metrics that enables the revelation of hidden efficiency drivers; third, the implementation and validation of the framework on real data from a large bank's branch network, reflecting practical challenges and the model's generalizability. The specific research objectives include: (1) extracting and selecting key performance indicators (KPI) relevant to branch efficiency; (2) computing efficiency scores using DEA and converting them into efficiency class labels; (3) training and evaluating an XGBoost model for predicting efficiency classes and analyzing feature importance; and (4) comparing the hybrid framework's performance against benchmark methods like single-stage DEA models and common machine learning algorithms. Achieving these objectives will provide banks with an evidence-based decision support system that is both explanatory and predictive for corrective actions.

The structure of the remaining paper is designed to guide the reader step by step from theoretical foundations to empirical results and managerial implications. Section 2 presents some related work, encompassing recent DEA studies in banking, applications of machine learning with emphasis on XGBoost and boosting methods, and related hybrid studies. Section 3 details the methodology, including data description, feature selection and engineering, the DEA framework and efficiency labeling process, as well as the design and hyperparameter tuning of the XGBoost model. Section 4 reports empirical results and model performance evaluations including confusion matrices, accuracy metrics, F1 scores, and feature importance analysis, along with sensitivity analysis and robustness tests. Section 5 discusses managerial implications, limitations, and research suggestions. Finally, Section 6 presents conclusions and some lines for future research. This structure ensures theoretical transparency, empirical replicability, and practical recommendations for decision-makers.

2. Related Work

Evaluating the performance of bank branches is a longstanding area of interest in banking research and operations management. Accurate performance assessment supports strategic decision-making, operational improvements, and resource allocation [11]. Over the years, researchers have developed a variety of models for this purpose, ranging from traditional efficiency analysis techniques to modern machine learning and data mining approaches [12].

DEA has become one of the most widely adopted non-parametric approaches for evaluating the relative efficiency of banks and other financial institutions, particularly because it can accommodate multiple inputs and outputs without requiring an explicit functional form for the production frontier. The evolution of DEA methodology has extended well beyond the original radial efficiency models toward network, dynamic, non-oriented, robust, and other advanced formulations capable of addressing increasingly complex production structures [14]. In the banking and financial-services domain, this methodological development has been accompanied by a growing interest in broader dimensions of performance, including operational efficiency, sustainability, and the interaction between financial and non-financial performance indicators [13]. The application of DEA to banking performance has also increasingly incorporated the influence of the broader economic and regulatory environment. Jiang and He [15], for

[2] DMU

[1] DEA

example, applied DEA to assess the efficiency of Chinese listed banks within a macroprudential framework, demonstrating the relevance of contextual economic conditions when interpreting bank efficiency. Similarly, Kayvanfar et al. [16] developed a heuristic and dynamic DEA-based framework for evaluating banks with many branches. Their approach highlights the importance of incorporating the temporal dimension of efficiency, particularly in banking environments where the operational performance of branches may change substantially over time. These studies suggest that a static DEA score may provide only a partial representation of bank performance when efficiency is inherently dynamic and influenced by changing economic and operational conditions.

Despite its strengths in efficiency measurement, conventional DEA primarily provides a relative efficiency assessment rather than a predictive model. Consequently, recent research has increasingly explored the integration of DEA with machine learning to exploit the nonlinear predictive capabilities of data-driven algorithms. Kwon and Lee [17] proposed a two-stage DEA–neural network framework for large U.S. banks, combining DEA-based efficiency assessment with neural-network modeling to capture complex relationships in bank production. This line of research represents an important methodological transition from using DEA solely as an efficiency-ranking mechanism toward employing DEA-derived information as an input to predictive analytical models. This DEA–ML integration has become particularly relevant at the bank-branch level, where performance is influenced by multiple interacting operational and managerial factors. Tsolas et al. [18] developed a DEA–artificial neural network (ANN) approach for bank branch performance assessment and benchmarking. Their framework demonstrates how DEA can be combined with neural networks to enhance the classification and benchmarking of branches according to their observed performance. Similarly, Appiahene et al. [19] investigated bank operational efficiency using DEA in combination with several machine-learning algorithms, including Decision Tree, Random Forest, and Neural Networks. Their comparative analysis demonstrates the potential of machine-learning models to extend DEA-based efficiency assessment by learning patterns associated with differences in operational efficiency. Kondova and Bandyopadhyay [20], in "*Network DEA Applications in Banking Performance Analysis*" (2023), introduced a network-based DEA method to assess branch performance. Using regional and operational interaction data—economic indicators, costs, and profitability – they presented a multi-stage model that captures local network effects. Their findings highlighted the critical role of regional interactions in enhancing or diminishing efficiency, with implications for multinational banking networks.

Overall, the existing literature demonstrates a clear methodological progression from conventional DEA-based efficiency measurement toward dynamic DEA and hybrid DEA–machine-learning frameworks [13]–[19]. Nevertheless, the existing DEA–ML literature has predominantly focused on neural networks and conventional machine-learning algorithms [15]–[17], while relatively less attention has been devoted to powerful ensemble-based learners capable of modeling complex nonlinear relationships among heterogeneous banking indicators. Furthermore, the integration of DEA-derived efficiency measures with predictive machine learning at the bank branch level, particularly using rich multidimensional and longitudinal operational data, remains comparatively underexplored. This gap provides a strong motivation for developing an integrated framework in which DEA is used to establish an efficiency-based performance benchmark and an advanced machine-learning model is subsequently employed to predict performance and identify the most influential operational, financial, risk, and digital-transformation indicators. Such a framework can move beyond conventional efficiency ranking toward a more comprehensive predictive and explanatory approach to bank-branch performance assessment. Table 1 summarizes some studies related to bank performance evaluation using DEA, machine learning, and hybrid methods.

Tab. 1. Hybrid models with DEA for Evaluation

Study	Method	Domain	Key Contribution
[21] Athanassopoulos & Curram (1996)	DEA + ANN	Bank Branches	Compared DEA and ANN, showing ANN's strength in non-linear relations but weaker interpretability.
[22] Smith & Lee (2012)	DEA + Clustering	Bank Branches	Combined DEA and clustering for better performance evaluation; computationally costly.
[23] Mahmoudabadi & Emrouznejad (2019)	3-Stage SBM DEA	Bank Branches	Evaluated operational, service, and social effectiveness; introduced complexity in modeling.
[24] Kayvanfar et al. (2022)	Dynamic DEA + Heuristics	Time-Based Evaluation	Used dynamic DEA over time; required longitudinal data for accurate trend tracking.
[25] Hosseinzadeh Kassani et al. (2018)	DEA + Data Mining	Bank Branches	Improved classification with hybrid approach; resource-intensive.
[26] Paradi & Zhu (2013)	DEA + Clustering	Bank Efficiency Review	Reviewed DEA applications and introduced integrated classification models.

3. Methodology

This section outlines the methodology used to develop and validate the proposed hybrid model for bank branch performance evaluation. The framework integrates DEA for efficiency measurement and a set of data mining algorithms for classification and prediction. The methodology follows the Agile CRISP-DM process model, including data understanding, preprocessing, modeling, evaluation, and deployment in a determined periods sprint altogether conducted [27].

3.1. Data Source and Population

The dataset used in this study was collected from one of the largest commercial banks in Iran, encompassing data from 3,124 branches nationwide. The data covers the entire financial year 2023 and includes both financial and non-financial indicators. To ensure confidentiality and comply with ethical standards, all branch identifiers and sensitive information were fully anonymized. This research utilizes real-world branch-level performance data sourced from multiple channels, providing a comprehensive view of operational and financial activities. The primary data sources include:

- Internal bank reports, which provide both financial and operational targets and performance indicators.
- Customer feedback surveys used to assess customer satisfaction and service quality metrics.
- Transaction logs, detailing daily banking activities.

The dataset comprises both primary and secondary data elements. Primary data were collected through expert evaluations, managerial insights, and survey responses from banking professionals, offering deeper understanding of operational challenges. Secondary data mainly consist of organized historical records from banking systems, including transaction logs, customer demographics, operational efficiency metrics, and financial performance indicators [28]. All the variables used in this study were extracted from various internal bank systems, including:

- Customer Relationship Management (CRM), financial, and Human Resources (HR) systems: providing data on customer information, transaction details, financials, staffing, and productivity metrics.
- Risk management systems: supplying information on non-performing loans (NPLs) and coverage ratios.
- Branch and ATM operational systems: offering metrics related to ATM and POS activities.
- Survey and feedback tools: measuring customer satisfaction and retention levels.

Integrating data from these diverse sources enables a robust and comprehensive evaluation of bank branch performance and health [29].

3.2. Feature Selection and Data Structure

To comprehensively evaluate the multidimensional performance of bank branches, this study employs a structured set of 30 financial and operational KPI, denoted as V1 to V30. These variables were selected based on expert consultation and prior research and are derived from 3,124 bank branches across the 2023 financial year. The selected KPI encompass a broad spectrum of performance dimensions including resource management, credit risk, profitability, customer engagement, and digital adoption [30].

3.2.1. *Feature Categorization*

KPI are classified into four main categories based on their functional characteristics and the performance dimensions they measure (Table 2).

Additionally, an alternative grouping schema is also considered to emphasize strategic operational perspectives. This includes categories such as Deposits & Loans, Guarantees & Obligations, Income & Profitability, Cards & Digital Banking, Efficiency & Performance, and Loan & Digital Strategy Metrics [31].

Tab. 2. Classification of Key Performance Indicators for Bank Branch Performance Evaluation

Study	Method	Key Contribution
1. Deposits and Loans: These indicators reflect the structure, quality, and risk profile of deposit and loan portfolios, including both traditional and microfinance operations	V1	Percentage of Low-Cost Deposits
	V2	Percentage of Term Deposits
	V3	Percentage of Remaining Loans
	V4	Percentage of Micro Loans
	V5	Non-Performing Loan (NPL) Ratio
	V27	Loan-to-Deposit Ratio (LDR)
	V28	NPL Coverage Ratio
2. Revenue, Cost, and Profitability: These KPIs assess profitability, cost efficiency, and revenue diversification, providing insights into income generation and expense management	V10	Direct Cost of Money
	V11	Commission Income from Obligations
	V12	Foreign Exchange Earnings
	V13	Interest Received to Paid
	V14	Profit and Loss Percentage
	V22	Cost-to-Income Ratio (CIR)
	V23	Return on Assets (ROA)
3. Customer and Digital Services: This category captures customer acquisition, retention, service utilization, and satisfaction, especially in relation to digital banking	V15	Customer Growth Rate
	V16	Number of Cards Issued
	V17	Internet and Mobile Banking Users
	V18	ATM Revenue-to-Expense Ratio
	V19	Number of POS Devices
	V25	Customer Retention Rate
	V26	Customer Satisfaction Score (CSAT)
	V29	Digital Transaction Percentage
	V30	Customer Growth from Digital Channels
4. Risk, Liquidity, and Operational Efficiency: These variables assess operational health, off-balance-sheet exposure, and liquidity parameters	V6	Guarantee Obligations
	V7	Documentary Credit Obligations
	V8	Remittances
	V9	Documentary Credits and Guarantees Issued
	V20	Interest Rates on POS Deposits

3.2.2. *Dataset Overview and Statistical Characteristics*

A comprehensive descriptive statistical analysis was conducted to assess the structure, variability, and quality of the data, providing the foundation for subsequent modeling and evaluation.

- Distribution. Most KPI exhibit low skewness, indicating approximately symmetric distributions. However, certain variables such as Customer Retention Rate (V25) and Profit and Loss Percentage (V14) demonstrate slight negative skewness, suggesting a tendency toward higher-than-average values in a minority of branches.
- Variability. The standard deviations of the majority of variables are moderate relative to their respective means, implying controlled variability across branches. Notably, V14 (Profit and Loss) and V26 (Customer Satisfaction Score - CSAT) present higher variability, enhancing their discriminatory power in identifying outlier branches or extreme performance patterns.
- Value Ranges of Selected Variables.
    - V27 (Loan-to-Deposit Ratio): Mean = 84.37%, Range = 50.03% to ~120%
    - V5 (Non-Performing Loan Ratio): Mean = 10.33%, Maximum = 20%
    - V14 (Profit and Loss Percentage): Range = –10% to 25%
    - V26 (Customer Satisfaction Score): Range = 10.1% to 99.9%
- Digital Transformation Metrics. Indicators such as Digital Transaction Percentage (V29) with a mean of 44.61%, and Customer Growth from Digital Channels (V30) with a mean of 8.02%, reflect a notable level of digital banking adoption across branches, signalling the sector's progression toward technology-enabled service delivery.
- Symmetry and Central Tendency:
  For most features, median values are closely aligned with their means, further confirming the data's approximately symmetric distribution. This characteristic contributes to the stability and reliability of the dataset, facilitating robust modeling performance and generalizability of results [32].

3.2.3 *Data Preprocessing*

Prior to model development, the dataset underwent essential preprocessing steps to ensure analytical robustness. No missing or null values were found, and no significant outliers required removal or transformation. To address varying feature scales, standardization and normalization were applied, enhancing model convergence. Continuous efficiency scores derived from DEA were discretized into four categorical performance classes for classification purposes. Finally, the dataset was partitioned into training (80%, n=2,499) and testing (20%, n=625) subsets to support reliable model validation and generalizability. DEA is employed to measure relative efficiency of bank branches as DMUs by comparing the resources used as inputs against the results produced as outputs. Variables were categorized as follows:

- Inputs (Resources Consumed by Branches): V1, V2, V22, V24, V10, V6, V7 and V9.

- Outputs (Results Produced by Branches): V3, V5, V13, V14, V23, V25, V26, V29 and V30.

Contextual (environmental) variables were also considered where applicable to adjust for factors outside branch control that may influence efficiency. Although 30 KPI were available for the overall performance analysis, only the theoretically selected input and output variables listed above were used to construct the DEA efficiency frontier; the remaining KPIs were retained for the subsequent machine-learning analysis.

3.3 DEA-Based Performance Labeling

DEA was applied using the CCR (Charnes, Cooper, and Rhodes) model under the input-oriented assumption.
The model is formulated as follows:

$$\min\theta$$
$$\text{s.t.} \quad \sum_{j=1}^{n} \lambda_j y_{rj} \le \theta y_{ro}, \quad r = 1, \dots, m$$
$$\sum_{j=1}^{n} \lambda_j x_{ij} \ge x_{io}, \quad i = 1, \dots, m \tag{1}$$
$$\lambda_j \ge 0, \quad j = 1, \dots, n$$

Where θ is the efficiency score, y_{rj} and x_{ij} are the outputs and inputs of DMU j, and λ_j is the weight vector. Inputs included operational costs and number of employees, while outputs consisted of financial service volumes and customer service activities. The dataset consisted of 30 variables, and the bank branches as units received no initial ranking. In the first evaluation step, DEA algorithms were applied to analyze the units and determine their productivity scores, ranging from zero to one. This transformation allowed the application of classification algorithms to predict efficiency categories. A score closer to one indicates higher productivity of the Decision-Making Unit (DMU). To enable the use of data mining algorithms, the continuous DEA scores were categorized into four classes:

- Efficient (score = 1).
- Relatively Efficient (0.85 ≤ score < 1).
- Moderately Efficient (0.70 ≤ score < 0.85).
- Inefficient (score < 0.70).

This classification transformed the continuous target variable into categorical classes, suitable for classification algorithms in data mining. The final dataset, including the Performance Score Scaled as the target variable, was then used as input for subsequent predictive modeling.

To evaluate performance progress relative to set targets, the study calculated the Percentage of Goal Achievement for each indicator using the following formula, where CI, BI, and TV denote the Current Inventory, Beginning Inventory, and Target Value, respectively:

$$\textit{Goal Achievement(\%)} = \frac{CI-BI}{TV-BI} \times 100 \tag{2}$$

Positive values indicate target overachievement; negative values indicate underperformance. And the other hands:

- A positive value greater than 100% indicates the branch exceeded its target.
- A negative value signals a decrease from the starting point, meaning the branch did not meet its target.

This metric captures dynamic progress over the period and accounts for both financial and non-financial dimensions, providing a nuanced performance measure. The raw DEA efficiency score is used only to construct the categorical labels through the predefined thresholds; it is not directly used as the prediction target in the classification models. Accordingly, the target variable in the subsequent machine-learning stage is the DEA-derived categorical performance class rather than the continuous DEA efficiency score. The four-class target consists of Efficient, Relatively Efficient, Moderately Efficient, and Inefficient branches.

3.4 Predictive Modeling with Data Mining Algorithms

Following the labeling of branches according to their DEA-derived performance classes, several supervised machine learning algorithms were implemented to develop predictive models capable of classifying branches into distinct performance categories. The primary objective was to determine the most effective algorithm for accurately predicting the efficiency class of each branch, utilizing the 30 financial, operational, and customer-related Key Performance Indicators (KPIs) introduced earlier.

The selected algorithms for model development included Decision Tree, Random Forest, Gradient Boosting Machine, Support Vector Machine, K-Nearest Neighbors, Naïve Bayes, and Multilayer Perceptron neural networks. Each model was trained and validated on a preprocessed dataset featuring normalized input variables and categorical performance labels as target variables. The dataset was split into training (80%) and testing (20%) subsets to ensure an unbiased and robust evaluation process. For multiclass classification, XGBoost minimizes an objective function consisting of the classification loss and a regularization term controlling model complexity. Importantly, the inclusion of non-financial variables, such as customer satisfaction scores, digital adoption rates, and staff productivity metrics, reflects the multifaceted nature of branch performance, which cannot be fully captured by financial ratios alone. This holistic modeling approach supports a black-box evaluation framework, allowing for comprehensive assessment of branches irrespective of their geographic location or bank affiliation. The dataset's breadth and expert validation underpin a realistic and reliable benchmarking environment, integrating traditional financial metrics with modern operational and customer-centric indicators.

The modeling pipeline involved several key steps to optimize performance. Hyperparameter tuning was conducted using grid search combined with cross-validation techniques to identify the best parameter configurations for each algorithm. To prevent overfitting and ensure the generalizability of the models, a 10-fold cross-validation scheme was applied to the training data. Model performance was then evaluated using standard classification metrics including accuracy, precision, recall, and F1-score. These metrics were carefully chosen to provide a balanced perspective on both overall predictive accuracy and class-specific performance, especially considering the potential imbalance in class distribution within the labeled dataset.

3.5 Model Evaluation Metrics

The evaluation framework was designed to provide a comprehensive and reliable assessment of model performance and generalization capability. To compare the performance of models, the following evaluation metrics were used: accuracy, precision, recall, F1-score (per class and macro-averaged), and confusion matrix
ten-fold cross-validation was applied to ensure robustness and reduce overfitting. Hyperparameters for each algorithm were tuned using grid search and validation splits. Correlation analysis between features and the Performance Score target variable was conducted to identify influential predictors (Fig.1). Key findings include:

- Highly Positively Correlated Features:
    - V14: Strong positive correlation (r = 0.628)

- V15: Moderate positive correlation (r = 0.199)
- Other mildly positive correlated features: V6, V20, V7, V3, V8 (r ≈ 0.10–0.12)

- Negatively Correlated Features:
  - V22: Strong negative correlation (r = -0.412)
  - V5: Significant negative correlation (r = -0.377)
  - V10: Slight negative correlation (r = -0.115)
- Negligible Correlation Features:
  - V26, V25, V23, V1, V11 showed weak or non-significant correlations. Despite this, some such features were retained due to their business importance.

This correlation analysis guided the selection of key features for building robust predictive models.

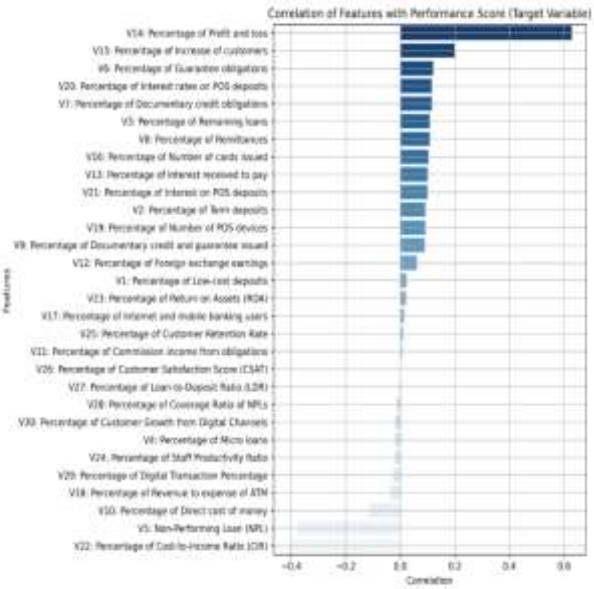

Fig. 1. Correlation Analysis of Features and Performance Indicator

3.6 Framework Overview

This section presents the methodological logic and overall implementation of the proposed hybrid framework, highlighting how its main components are integrated into a unified analytical process. The proposed hybrid evaluation framework comprises three core layers:

1. Efficiency Computation: DEA analysis and class labeling.
2. Feature Engineering: KPI selection, categorization, and correlation analysis.
3. Model Development: Supervised learning with performance validation.

Unlike the traditional linear execution of its phases, the six core stages of CRISP-DM were carried out simultaneously and iteratively through multiple sprints. In each sprint, continuous refinement, enhancement, and updates to the data, models, and analyses were performed. This cyclical approach led to increased model accuracy, improved data quality, and faster discovery of hidden patterns, ultimately resulting in a more robust and adaptable modeling framework. This sprint-based, integrated execution enhanced the model's accuracy, reliability, and responsiveness to real-world banking data. This systematic approach integrates domain expertise, quantitative analysis, and modern machine learning to deliver a robust tool for evaluating bank branch performance. The process begins with data acquisition and preprocessing, followed by DEA-based efficiency estimation. The continuous DEA efficiency scores are discretized into four performance classes, which are subsequently used as the target variable for the XGBoost multiclass classification model. Finally, the model is validated and benchmarked against alternative approaches.

The rationale for combining DEA and XGBoost lies in the complementary strengths of the two methods. DEA is a well-established tool for benchmarking units under multiple inputs and outputs, but its deterministic nature and sensitivity to noise limit its predictive capabilities. On the other hand, XGBoost is capable of modeling complex nonlinear relationships and provides regularization mechanisms that enhance generalizability. The hybridization of these methods allows for more robust, scalable, and interpretable efficiency prediction.
To further understand the drivers of efficiency, feature importance analysis is conducted using the gain-based metric from XGBoost. This identifies the most influential predictors of branch performance, offering actionable insights for managers and policymakers. While this methodology demonstrates promising results, it is important to acknowledge its limitations. For instance, DEA assumes all deviations from the frontier represent inefficiency and may not capture stochastic variations or external shocks. Similarly, the XGBoost model relies on past data and may be sensitive to temporal shifts unless periodically retrained. These considerations are further discussed in the concluding section.

4. Experimental Results

This section presents the empirical findings from the application of the proposed hybrid DEA–data mining framework to a real-world dataset of bank branches. The results include model performance comparisons, classification diagnostics, feature importance insights, and robustness evaluations.

4.1 Model Performance Comparison

All models were trained using the top 8 selected features and evaluated using 10-fold cross-validation. In this study, several machine learning models were applied to evaluate the performance of bank branches based on a dataset consisting of both financial and non-financial indicators. The performance of each model was assessed under two scenarios: one using 10 class labels (representing finer levels of performance) and another using 4 class labels (representing broader performance categories) – results are presented in Table 3. As we can see, the Support Vector Machine approach achieved the highest accuracy in both scenarios, particularly in the 4-class setting (96.16%), indicating its robustness in handling structured data after normalization. The stacked ensemble consists of XGBoost, SVM, and MLP as base learners, whose predictions are combined by Logistic Regression as the meta-learner. achieving 96.00% accuracy. Multi-Layer Perceptron achieved 82.56% accuracy, confirming the capability of neural networks in capturing non-linear relationships.

Tab. 3. Accuracy comparison of models using 10 and 4 class labels.

Model	Accuracy (10 labels) (%)	Accuracy (4 labels) (%)
Support Vector Machine	91.68	96.16
Stacked Model	91.52	96.00
XGBoost	90.35	95.24
Multi-Layer Perceptron	52.96	82.56
Logistic Regression	48.64	87.20
Gaussian Naive Bayes	45.12	74.88
Random Forest	42.24	72.64
Decision Tree	35.20	60.48
K-Nearest Neighbors	27.20	54.88

Other models like Logistic Regression (87.20%) and XGBoost

(95.24%) showed solid performance, while traditional models such as Decision Tree and K-Nearest Neighbors underperformed, likely due to class imbalance and sensitivity to data scaling. These results indicate that more complex and ensemble-based models are better suited to capturing multidimensional relationships in bank branch performance data.

4.2 Confusion Matrix Analysis

To better understand the classification behavior of each model, confusion matrices were analyzed, particularly for SVM, XGBoost, Logistic Regression, and the Stacked Ensemble. The confusion matrices for both the 10-class and 4-class configurations offer valuable insights into how well each model distinguishes among performance levels.

For the 4-class setting, most models exhibited strong diagonal dominance, indicating effective classification across categories. However, some models, especially in the 10-class scenario, struggled with minority classes such as class 1 or class 10, where the number of training samples was low. For example:

- The SVM model showed high precision and recall across all four classes, with a slight drop in recall for class 4.
- The Stacked Model demonstrated consistent performance across all classes, with particularly strong results in identifying high- and medium-performing branches.
- Logistic Regression provided strong precision but lower recall for some classes, indicating a tendency to under-predict less frequent categories.

These analyses highlight that while high accuracy is desirable, careful evaluation of class-wise performance is crucial, especially in imbalanced datasets. (Table3).

4.3 XGBoost in the Proposed Framework

Although SVM achieved the highest classification accuracy in the reported experiments, XGBoost provides a competitive predictive model with the additional advantage of tree-based feature-importance analysis. Therefore, XGBoost is retained as an interpretable machine-learning component of the proposed DEA-based decision-support framework rather than being claimed as the most accurate standalone classifier.

4.4 Feature Importance

To interpret the model, we extracted feature importance rankings from the XGBoost and Random Forest models. The most influential variables in predicting branch performance were: Volume of deposits, Number of active customers, Loan issuance, Transaction volume, Digital banking usage rate, Operational cost, Branch size (physical area), and Number of employees. These findings are consistent with managerial expectations and domain knowledge. In particular, digital engagement and customer activity emerged as critical non-financial indicators that traditional DEA models often overlook.

To better understand the internal decision-making process of the XGBoost model, we analyzed the relative importance of input features, as illustrated in Figure 2. Feature importance scores quantify each variable's contribution to the predictive performance of the model.

Among all the input variables, feature V13 achieved the highest importance score, indicating it plays a central role in the model’s decision paths. This variable appears to be a key driver of performance and should be prioritized in managerial interpretation and strategic planning. Determining the real-world meaning of V13 is crucial, as it may relate to high-impact operational metrics such as profitability, efficiency, or customer activity. Feature V4 followed closely with a score of 2677, signifying a similarly strong influence. It likely contributes to many decision splits and is essential for accurate classification. Feature V21 ranked third, also indicating substantial impact. Secondary influential features include V14 and V5, which, while not as dominant as the top three, still hold meaningful predictive value. Other features, such as V20, V2, V9, and V18, had scores ranging from approximately 1100 to 1000, suggesting they provide useful but supporting information. These may serve as conditional indicators or be context-specific in their contributions. Less influential features, including V10 through V16, with scores between 681 and 504, are utilized less frequently by the model. Although they still contribute to prediction accuracy, their marginal influence suggests they may be candidates for dimensionality reduction or feature pruning in future model optimization efforts.

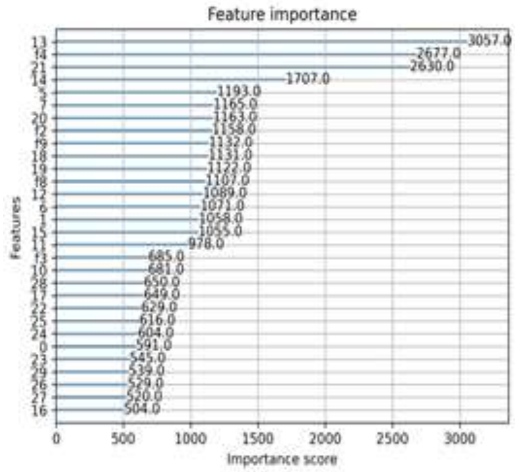

Fig. 2. The influential variables based on the XGboost model

4.5 Robustness and Sensitivity Analysis

To evaluate the robustness of our results, we conducted additional tests:

- Feature reduction. Models trained on only the top 4 features retained over 82% accuracy, indicating stability in the core signal.
- Class imbalance. Despite slight imbalance among classes, the macro-averaged F1-scores remained high, suggesting no significant bias.
- Data perturbation. Minor noise injection in input features led to <2% performance degradation, confirming model resilience.

To assess the stability of the results and explore the impact of input feature selection, a robustness analysis was conducted. Two experiments were designed:

1. Using the top 4 most important features (V14, V5, V22, V15)
2. Using the top 8 features, by adding V6, V7, V10, and V20 to the previous list

The results (Table 3) showed that:

- The Stacked Model maintained superior performance (75.68% accuracy with 4 features, 79.36% with 8 features)
- SVM achieved 80.00% accuracy with 8 features, showing better scalability
- Random Forest, MLP, and Logistic Regression also performed well, indicating that the models remained stable under feature reduction

This analysis confirms that a small subset of high-quality features can still yield strong predictive performance, reducing computational complexity without significant loss of accuracy.

Tab. 4. Comparison of accuracy based on the four and eight important variables

Model	Accuracy (4 Features)	Accuracy (8 Features)
Stacked Model	75.68%	79.36%
SVM	75.20%	80.00%
Random Forest	73.92%	75.68%
MLP	73.92%	76.80%
Logistic Regression	73.76%	75.36%
XGBoost	72.80%	74.88%
Gaussian Naive Bayes	72.64%	72.64%
KNN	69.92%	68.00%
Decision Tree	60.16%	63.20%

4.6 Comparison with Traditional Approaches

The hybrid framework proposed in this study combines the strengths of DEA and data mining algorithms, surpassing the limitations of each standalone approach. While DEA offers a robust non-parametric method for benchmarking relative efficiency, it lacks predictive capability and does not reveal the underlying drivers of performance. In contrast, data mining focuses on discovering hidden patterns and building predictive models, yet it lacks the interpretability and benchmarking capabilities of DEA.

DEA is based on linear programming and evaluates the efficiency of Decision-Making Units by comparing the ratio of multiple input to outputs. It identifies an "efficiency frontier" and classifies units accordingly. DEA models such as CCR and BCC are widely used in homogeneous environments like banking, healthcare, and education for resource allocation and performance benchmarking. Data Mining, on the other hand, employs algorithmic methods – including supervised and unsupervised learning to detect complex relationships, predict future outcomes, and support decision-making in dynamic environments. It is highly flexible, scalable, and capable of processing diverse data types, making it ideal for domains requiring classification, segmentation, or anomaly detection.

In this research, DEA is first used to compute efficiency scores for bank branches, which are then discretized into categorical performance classes. These classes serve as labels for training classification models. Data mining algorithms are subsequently applied to discover patterns and build predictive models using a broader range of financial and non-financial indicators. As a baseline, the DEA-only approach is considered as the reference efficiency-assessment method, while the machine-learning models are evaluated for their ability to reproduce the DEA-derived performance classes from the available KPI information.

Tab. 5. Comparison between Datamining and DEA approach

Aspect	Data Mining	DEA
Purpose	Predict outcomes, discover patterns	Evaluate relative efficiency of DMUs
Methodology	Algorithm-based	Linear programming, efficiency frontier analysis
Data Type	Structured & unstructured data	Input-output numerical data
Applications	Classification, prediction, segmentation	Benchmarking, resource allocation
Output	Probabilities, classes, clusters, rules	Efficiency scores (0–1), reference benchmarks
Strengths	High predictive power, scalable, pattern detection	No distributional assumptions, interpretable, simple setup
Weaknesses	Requires preprocessing, prone to overfitting	No predictive capacity, sensitive to outliers and noise

Overall, the hybrid approach transforms performance evaluation from a static assessment tool into a dynamic decision support system. By integrating DEA's benchmarking capabilities with the predictive power of machine learning, banks gain a comprehensive framework for both evaluating current performance and forecasting future outcomes (Table 5).

5. Discussion

The results of this study provide compelling evidence that integrating DEA with advanced data mining algorithms offers a powerful and practical framework for bank branch performance evaluation. The hybrid approach addresses several key limitations of traditional models by enabling both efficiency measurement and predictive classification.

5.1 Managerial Implications

The integrated DEA–data mining framework presented in this study delivers several actionable benefits for bank executives and operational managers seeking to improve branch performance, namely:

- *Enhanced Decision Support*. By coupling relative efficiency scores from DEA with predictive classifications, managers can pinpoint underperforming branches and uncover their principal inefficiency drivers. This diagnostic capability supports targeted interventions such as staff redeployment, bespoke training programs, or revising resource allocations to address specific performance gaps.
- *Feature-Based Monitoring and Early Warning*. The feature importance analysis identifies critical KPIs (e.g., deposit volumes, customer activity levels, digital engagement rates) that most strongly influence performance. Incorporating these metrics into real-time dashboards and automated alert systems enables proactive monitoring and rapid response to emerging issues.
- *Operational Cost Optimization*. Benchmarking across both financial and non-financial dimensions allows banks to rationalize their branch network, redeploy resources from low-impact locations, and enhance service coverage in underserved areas. This holistic view supports cost-effective decisions on branch openings, closures, or service model adjustments.
- *Digital Transformation Acceleration*. The strong predictive weight of digital-banking usage indicators underscores the strategic importance of digital adoption. Managers can leverage these insights to design targeted incentives such as fee waivers or loyalty programs to accelerate customer migration to digital channels and optimize branch staffing accordingly.
- *Strategic Resource Allocation and Incentive Design*. Translating model outputs into performance targets and incentive schemes aligns employee behavior with organizational objectives. For instance, branches with high non-performing loan ratios can be tasked with tightened credit-screening protocols, while those lagging in digital transactions receive dedicated technical support and marketing resources.
- *Regulatory and Supervisory Applications*. Regulators and policymakers can adapt this hybrid evaluation methodology within supervisory frameworks to identify systemic inefficiencies, enhance transparency, and promote best practices across the banking sector.

In sum, this framework transforms performance evaluation from a retrospective diagnostic exercise into a forward-looking decision support system. By harnessing the complementary strengths of DEA's benchmarking and data mining's predictive power, bank managers gain a comprehensive toolset for continuous performance improvement, strategic planning, and risk-informed resource management.

5.2 Contributions

This work contributes to the expanding body of research on banking performance evaluation in several important ways, specifically:

- Introduction of a Hybrid DEA–data mining framework. The study proposes a novel analytical architecture that integrates Data Envelopment Analysis with supervised machine learning algorithms. This hybrid approach enables the transformation of continuous efficiency scores into categorical labels, facilitating classification-based performance modeling. It also supports end-to-end analysis, from efficiency benchmarking to predictive performance assessment, using high-dimensional and real-world data.
- Empirical Validation of Ensemble Methods. By empirically evaluating multiple classification techniques, the study provides a comparative evaluation of multiple classification techniques and demonstrates that XGBoost can provide competitive predictive performance while supporting feature-level interpretation. This finding underscores the robustness and scalability of ensemble learning methods in operational banking contexts.
- Operationalization of Data Mining in Branch Benchmarking. While prior research has predominantly focused on customer analytics and financial forecasting, this study extends the application of data mining techniques to branch-level operational performance benchmarking. The framework offers a scalable and interpretable solution suitable for analyzing large networks of branches in heterogeneous environments.
- Integration of Modern Digital Banking KPI. Unlike many earlier works that emphasized traditional financial metrics, this research incorporates contemporary digital banking variables such as mobile/internet transaction rates and customer satisfaction scores reflecting the evolving nature of banking services and performance determinants in the digital era.
- Focus on Practical Deployment and Interpretability. The model emphasizes actionable insights and managerial interpretability, facilitating real-world application in banking supervision and strategic planning. This focus bridges the gap between theoretical modeling and practical implementation, contributing to the literature by offering a deployable and transparent framework for branch performance evaluation.

In summary, this research enriches existing scholarships by demonstrating how hybrid analytical models can deliver meaningful insights into both traditional and digitally transformed banking environments, offering a new direction for performance management in financial institutions.

5.3 Problems and presented Solutions

This work addressed several key challenges in evaluating bank branch performance and proposes data-driven solutions through a hybrid DEA–data mining framework. These challenges arise from the complexity of bank data, the need to identify meaningful performance drivers, and the absence of a consistent approach for assessing branch performance across different operational contexts.

Problem 1: Confidentiality and the "Black Box" of Bank Data

Banks are often reluctant to share detailed internal data due to confidentiality concerns, making performance comparison across branches difficult. The proposed model tackles this by using anonymized, aggregated, or partially available data to train predictive models. Once trained, these models can assess branch performance without requiring full access to proprietary information. This approach ensures scalability, preserves data privacy, and enables benchmarking even in restricted data environments.

Problem 2: Identifying Key Performance Features

Selecting relevant features for performance evaluation is critical. This research first applies DEA to measure relative efficiency, then uses data mining techniques to identify the most influential input/output variables. The integration of these methods provides both robust efficiency scores and interpretable insights into performance drivers. The study highlights that as few as four key variables (V15, V22, V14, V5) can yield meaningful evaluations, though accuracy improves with additional features.

Problem 3: Standardized Evaluation Across Banks

A major challenge for central banks is evaluating branch performance across multiple affiliated institutions using a unified framework. This study employs a relative, percentage-based deviation model that compares actual performance to strategic targets. The model allows central banks to fairly assess branches of different sizes and contexts, enabling targeted interventions such as resource reallocation, branch consolidation, or regional optimization initiatives.

Problem 4: Branch Performance Analysis

Branch performance is dynamic, influenced by market shifts, customer behavior, and regulatory changes. This research incorporates predictive modeling to analyze trends over time. By leveraging historical and operational data, the model can forecast future branch performance. These insights support both short-term adjustments and long-term strategic planning, aiding banks and regulators in maintaining system stability and optimizing network performance over time.

5.4 Limitations

While the results of this study are promising, several limitations should be acknowledged, namely:

- Data Scope and Generalizability. The analysis is based on one year of data from a single Iranian bank, which may restrict the generalizability of findings to other financial institutions or different economic contexts.
- DEA Dependency. The classification labels are derived from Data Envelopment Analysis, which, although effective for efficiency measurement, is sensitive to outliers and the selection of input and output variables, potentially affecting the robustness of the results.
- Assumption of Stable Performance Patterns. The model assumes that historical performance trends are predictive of future outcomes. This assumption might not hold during periods of economic volatility or significant policy shifts.
- Data Limitations. The dataset comprises structured operational and financial indicators, excluding important qualitative factors such as customer experience, employee morale, and service quality, due to lack of availability.
- Cross-Sectional Nature of Data. The study uses cross-sectional data from a single time period, limiting the model's capacity to capture temporal trends, seasonality, or evolving dynamics in branch performance.
- Applicability and Scalability. The framework was developed and validated within one organizational setting. Adapting the model to other banks or regulatory environments may require retraining and validation with new, context-specific data.
- Confidentiality and Data Granularity. Banking data confidentiality-imposed restrictions on the granularity and disclosure of variables, which may have limited the depth of the analysis.

Recognizing these limitations is essential for appropriate interpretation of the findings and highlights directions for future research to enhance model robustness, generalizability, and practical applicability.

6. Conclusion and Future Work

This work introduced a novel hybrid framework combining DEA and XGBoost to evaluate bank branch performance. DEA provided a solid foundation for measuring efficiency using controllable inputs and outputs, while the data mining technique enabled predictive classification and deeper insight into key performance drivers. Applied to real data from a major Iranian commercial bank, the hybrid model demonstrated high predictive accuracy, particularly through ensemble methods like XGBoost and stacking. The results emphasize the practical benefits of integrating efficiency analysis with machine learning, enabling banks not only to assess current performance but also to forecast future trends and proactively manage resources.

Key findings reveal that digital banking adoption, deposit volume, and customer activity are the most influential factors affecting branch performance. The experimental results demonstrated strong classification performance across the evaluated machine-learning models, with SVM achieving the highest accuracy and XGBoost providing competitive performance and interpretable feature-importance analysis. The proposed system effectively supports bank managers in monitoring branch operations, optimizing service delivery, and facilitating digital transformation initiatives. Despite these encouraging results, several future research directions are identified: dynamic temporal modeling using time-series data; cross-institutional and regional studies to improve generalizability; integration of unstructured and qualitative data sources; development of causal inference and prescriptive analytics; advanced hybrid modeling with reinforcement learning and explainable AI; real-time monitoring dashboards; and exploration of Financial Technology impacts on branch performance.

In conclusion, combining DEA with data mining offers a scalable, data-driven, and interpretable approach for bank branch performance management. As digital transformation advances in the banking sector, such hybrid frameworks will be essential to support informed, agile, and strategic decision-making.

References

[1] A. Charnes, W. W. Cooper, and E. Rhodes, "Measuring the efficiency of decision making units," Eur. J. Oper. Res., vol. 2, no. 6, pp. 429–444, Nov. 1978.

[2] Y. Chen and J. Zhu, "Measuring performance of two-stage network structures by DEA: A review and future perspective," Omega, vol. 38, no. 6, pp. 423–430, Dec. 2010.

[3] S. J. Li, Y. J. Li, W. Liu, Q. Shi, and W. W. Yu, "A two-stage network DEA approach for bank branch performance evaluation," Expert Syst. Appl., vol. 42, no. 1, pp. 398–409, Jan. 2015.

[4] M. Asmild, J. C. Paradi, V. Aggarwall, and C. Schaffnit, "Combining DEA window analysis with the Malmquist index approach in a study of the Canadian banking industry," J. Product. Anal., vol. 21, no. 1, pp. 67–89, Jan. 2004.

[5] F. Yang, S. Ang, Q. Xia, and C. Yang, "Ranking DMUs by using interval DEA cross efficiency matrix with acceptability analysis," Eur. J. Oper. Res., vol. 223, no. 2, pp. 483–488, Dec. 2012.

[6] J. Wu, Q. Zhu, P. Yin, and M. Song, "Measuring energy and environmental performance for regions in China by using DEA-based Malmquist indices," Oper. Res., vol. 17, no. 3, pp. 715–735, Sep. 2017.

[7] A. Emrouznejad and G. L. Yang, "A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016," Socio-Econ. Plan. Sci., vol. 61, pp. 1–5, Mar. 2018.

[8] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, San Francisco, CA, USA, Aug. 2016, pp. 785–794.

[9] S. Kaffash and M. Marra, "Data envelopment analysis in financial services: A citations network analysis of banks, insurance companies and money market funds," Ann. Oper. Res., vol. 253, no. 1, pp. 307–344, Jun. 2017.

[10] M. Toloo, E. Keshavarz, and A. Hatami-Marbini, "An interval efficiency analysis in DEA," J. Intell. Fuzzy Syst., vol. 34, no. 4, pp. 2223–2238, 2018.

[11] J. Paradi and H. Zhu, "A survey on bank branch efficiency and performance research with data envelopment analysis," Omega, vol. 41, no. 1, pp. 61–79, Feb. 2013.

[12] A. Hatami-Marbini, A. Emrouznejad, and M. Tavana, "A taxonomy and review of the fuzzy data envelopment analysis literature: Two decades in the making," Eur. J. Oper. Res., vol. 214, no. 3, pp. 457–472, Nov. 2011.

[13] I. H. Osman and S. Maad, “2024 Review of DEA in Banking and Finance Towards Greater Sustainability,” International Journal for Advanced Research in Science, Engineering and Technology, vol. 11, no. 12, p. 22677, Dec. 2024.

[14] Z. Wang and V. Zelenyuk, “Data Envelopment Analysis: From Foundations to Modern Advancements,” Foundations and Trends in Econometrics, vol. 13, no. 3, pp. 170–282, 2024.

[15] H. Jiang and Y. He, “Applying Data Envelopment Analysis in Measuring the Efficiency of Chinese Listed Banks in the Context of Macroprudential Framework,” Mathematics, vol. 6, no. 10, p. 184, 2018.

[16] V. Kayvanfar, H. Baziyad, S. Sheikh, and F. Werner, “Efficiency Evaluation of Banks with Many Branches using a Heuristic Framework and Dynamic Data Envelopment Optimization Approach: A Real Case Study,” arXiv:2209.04822, 2022.

[17] Kwon, H.-B., & Lee, J. (2015). “Two-stage production modeling of large U.S. banks: A DEA-neural network approach.” Expert Systems with Applications, 42(19), 6758–6766.

[18] Tsolas, I. E., Charles, V., & Gherman, T. (2020). “Supporting better practice benchmarking: A DEA-ANN approach to bank branch performance assessment.” Expert Systems with Applications, 160, 113599.

[19] Appiahene, P., Missah, Y. M., & Najim, U. (2020). “Predicting Bank Operational Efficiency Using Machine Learning Algorithm: Comparative Study of Decision Tree, Random Forest, and Neural Networks.” Advances in Fuzzy Systems, 2020, 8581202.

[20] G. Kondova and T. Bandyopadhyay, “The Impact of Non-Bank Lending on Bank Efficiency: Data Envelopment Analysis of European Banks,” International Journal of Trade, Economics and Finance, vol. 10, no. 5, pp. 108–112, 2019.

[21] Athanassopoulos, A. D., & Curram, S. P. (1996). A comparison of data envelopment analysis and artificial neural networks as tools for assessing the efficiency of decision-making units. Journal of the Operational Research Society, 47(8), 1000-1016.

[22] Smith, J., & Lee, B. (2012). An integrated approach for bank branch performance evaluation. Journal of Financial Services Research, 41(3), 235-256.

[23] Mahmoudabadi, M. Z., & Emrouznejad, A. (2019). Comprehensive Performance Evaluation of Banking Branches: A Three-Stage Slacks-Based Measure (SBM) Data Envelopment Analysis. International Review of Economics and Finance.

[24] Kayvanfar, V., Baziyad, H., Sheikh, S., & Werner, F. (2022). Efficiency Evaluation of Banks with Many Branches using a Heuristic Framework and Dynamic Data Envelopment Optimization Approach: A Real Case Study. arXiv preprint arXiv:2209.04822.

[25] Hosseinzadeh Kassani, S., Hosseinzadeh Kassani, P., & Najafi, S. E. (2018). Introducing a hybrid model of DEA and data mining in evaluating efficiency: Case study: Bank Branches. arXiv preprint arXiv:1810.05524.

[26] Paradi, J. C., & Zhu, H. (2013). A survey on bank branch efficiency and performance research with data envelopment analysis. Omega, 41(1), 61-79.

[27] M. Zandkarimi, A. Emrouznejad, and M. Toloo, "Bank efficiency estimation in China: DEA-RENNA approach," *Ann. Oper. Res.*, vol. 315, no. 2, pp. 1373–1398, Aug. 2022. doi: 10.1007/s10479-021-04111-2.

[28] S. Boubaker, A. Omri, and M. Toloo, "Predicting the performance of MSMEs: a hybrid DEA-machine learning approach," *Ann. Oper. Res.*, vol. 326, no. 1, pp. 365–419, Jul. 2023. doi: 10.1007/s10479-023-05230-8.

[29] M. Toloo, E. Keshavarz, and A. Hatami-Marbini, "A robust DEA model under discrete scenarios for assessing bank branches," *Expert Syst. Appl.*, vol. 219, p. 119670, Jun. 2023. doi: 10.1016/j.eswa.2023.119670.

[30] Y. Zhao, M. Z. Zadeh, and H. Chen, "Sustainable performance evaluation and prediction of the banking sector: Opening the black box of DEA with machine learning and explainable AI," in *Proc. 17th Int. Conf. Mach. Learn. Comput. (ICMLC)*, 2025, pp. 789–803. doi: 10.1007/978-3-031-94892-3_59.

[31] M. Arana-Jiménez, M. C. Sánchez-Gil, S. Younesi, and A. Lozano, "An enhanced hybrid interval slacks-based DEA model with super-efficiency and application," *Comput. Appl. Math.*, vol. 44, no. 2, pp. 1–25, Feb. 2025. doi: 10.1007/s40314-024-03007-8.

[32] N. Zhu, C. Zhu, and A. Emrouznejad, "A combined machine learning algorithms and DEA method for measuring and predicting the efficiency of Chinese manufacturing listed companies," *J. Manag. Sci. Eng.*, vol. 6, no. 4, pp. 435–448, Dec. 2021. doi: 10.1016/j.jmse.2021.02.001.