ISSN 2723-6471

Forecasting Bank Efficiency Using Data Envelopment Analysis with Directional Distance Functions and Machine Learning: Time-Series Validation and Shapley Value Interpretation

Chau Dinh Linh[1,*]

[1]*Ho Chi Minh University of Banking, 36 Ton That Dam Street, Sai Gon Ward, Ho Chi Minh City and 70000, VietNam*

(Received: November 20, 2025; Revised: January 20, 2026; Accepted: March 25, 2026; Available online: April 4, 2026)

Abstract

This study develops a structured framework to forecast the operational efficiency of commercial banks in Vietnam. The analysis is based on a balanced panel of 27 banks over the period 2016–2024. Bank efficiency is first measured using a directional distance function within a data envelopment analysis framework (DEA – DDF). This approach incorporates both desirable outputs and undesirable outputs related to credit risk. The estimated efficiency scores are then used as prediction targets in several machine learning models. Model performance is evaluated under both conventional test settings and time-series cross-validation, and predictions are interpreted using Shapley value–based analysis (SHAP). Under a conventional test set, the gradient boosting model (XGBoost) shows the best performance, with a root mean squared error of 0.060 and a coefficient of determination (R^2) of 0.353. However, when time-series cross-validation is applied to reflect realistic forecasting conditions, predictive accuracy declines sharply. The average coefficient of determination falls to approximately 0.005. This suggests that static validation can overstate performance and that forecasting efficiency in a changing financial environment remains difficult. The interpretation results provide additional insights. Net interest margin has a positive effect on predicted efficiency, although the effect weakens at very high levels. The cost-to-income ratio shows a threshold around 0.55, beyond which efficiency declines more strongly. Bank size has a largely neutral impact. The interaction between capital adequacy and profitability shows a conditionally negative pattern. Prediction errors are larger in the most recent year and among banks with very high efficiency scores. In summary, the results highlight both the potential and the limitations of machine learning in forecasting efficiency and emphasize the importance of time-aware validation.

Keywords: Bank Efficiency, DEA – DDF, Machine Learning, Time-Series Validation, SHAP.

1. Introduction

Commercial banks play a central role in Vietnam's financial system. Their operational efficiency reflects internal management quality and affects the stability of the broader system [1], [2]. For this reason, understanding and forecasting bank efficiency remains an important task. Forecasting efficiency, however, is not simple. Traditional econometric approaches, such as panel regressions and generalized additive models, are widely used because they are transparent and allow researchers to control for bank-specific effects. These models are useful for testing economic relationships. At the same time, they require researchers to specify the functional form in advance. When relationships between financial indicators and efficiency are highly non-linear or involve multiple interactions, such specifications may not capture all relevant patterns.

Machine learning models can handle complex and high-dimensional data more flexibly. However, their internal logic is often difficult to interpret, especially in regulated sectors such as banking [3]. In this study, SHAP is used to make model predictions more transparent. SHAP is not intended to replace traditional econometric methods. Instead, it serves as a complementary tool that helps explain complex predictive models.

This study follows four main steps. First, bank efficiency is measured using Data Envelopment Analysis with a Directional Distance Function, which accounts for risk-adjusted performance. Second, the estimated efficiency scores are used as dependent variables in machine learning models. The objective at this stage is predictive rather than causal.

***Corresponding author: Chau Dinh Linh (linhcd@hub.edu.vn)**
DOI: https://doi.org/10.47738/jads.v7i2.1244

Third, model performance is evaluated using time-series cross-validation to reflect realistic forecasting conditions and avoid information leakage. Finally, SHAP is applied to interpret the drivers of the model's predictions. Using data from 27 Vietnamese commercial banks over the period 2016–2024, the study makes three main contributions. First, it connects efficiency measurement, forecasting, validation, and interpretation within a single analytical structure. Second, the empirical results show that gradient boosting performs well under static validation but becomes more sensitive when temporal shifts are considered. Third, the SHAP analysis highlights threshold effects and non-linear relationships in profitability and cost efficiency that are difficult to detect using conventional regressions. These results may help managers and regulators better understand efficiency dynamics.

2. Literature Review

2.1. Bank Efficiency

Bank efficiency is a critical aspect of the financial industry, extensively researched to assess the operational performance and management capabilities of banking institutions [1], [4]. Efficient banks play a pivotal role in maintaining financial system stability [2], [5], while inefficiency can pose significant risks [2], [4]. Measuring bank efficiency often involves comparing banks' performance against an optimal efficiency frontier [5]. Key methods include traditional econometric (parametric) approaches and non-parametric approaches like Data Envelopment Analysis [6], [7]. Some studies have linked efficiency to risk management capabilities and bank survival [4]. However, academic debates also exist regarding the relationship between efficiency, competition, and stability, with some research suggesting that increased competition might boost efficiency but also potentially increase the fragility of the banking system [2].

2.2. Data Envelopment Analysis and Direction Distance Function Models

The Data Envelopment Analysis is a non-parametric mathematical method widely used to evaluate the relative efficiency of decision-making units, such as banks, by constructing an efficiency frontier from input and output data [8], [9]. A prominent advantage of DEA is its ability to concurrently handle multiple inputs and outputs without requiring assumptions about the functional form of their relationships, making it a particularly suitable tool for assessing the complex performance of banks [8], [10]. Basic DEA models include: (i) CCR Model: Assumes constant returns to scale [10]; (ii) BCC Model: Extends the CCR model to account for variable returns to scale, allowing the decomposition of technical efficiency into pure technical efficiency and scale efficiency [10].

The Directional Distance Function represents a significant advancement in DEA, enabling efficiency measurement by simultaneously expanding desirable outputs and contracting undesirable outputs in a predefined direction [11], [12]. DDF offers greater flexibility in analyzing various strategies that DMUs can employ to reach the efficiency frontier [13]. It can also handle negative input or output values and provides useful efficiency measures when input- and output-oriented technologies are unsuitable in non-competitive markets [14]. DDF further allows the decomposition of profit efficiency into directional technical and allocative efficiency [14]. Some studies have proposed enhanced DEA models based on DDF modifications to decompose technical efficiency into operational efficiency and risk management efficiency in the banking sector [15].

Application of DEA and DDF in Vietnam: Many studies have applied DEA to evaluate the efficiency of Vietnamese commercial banks. One study examined the period from 2008 to 2015 and reported relatively high and stable efficiency levels, with limited impact from the global financial crisis [16]. Other research compared parametric and non-parametric methods and found gradual improvements in overall efficiency [17], [18]. DEA window analysis has also been used to assess performance over time [19].

In addition to DEA, some studies have used SFA (Stochastic Frontier Analysis) and panel regression models. These studies provide useful evidence on differences across banks. However, most analyses rely on static models and linear assumptions. Validation is often conducted within a single estimation setting. The time-ordered nature of financial data is rarely addressed directly. Interpretation usually focuses on regression coefficients or average effects. Non-linear patterns and interaction effects receive limited attention.

ISSN 2723-6471

This study follows a different path. It combines DEA–DDF measurement, time-series forecasting, robustness checks, and structured interpretation within one framework.

2.3. Machine Learning in Efficiency Forecasting

Machine Learning (ML) has emerged as a powerful tool for forecasting in finance [20], [21]. ML algorithms can process large datasets, capture complex non-linear relationships, and generate accurate predictions [22]. In banking, ML is commonly applied to credit default prediction, fraud detection, and risk forecasting [23], [24].

In bank efficiency forecasting, ML models are used to predict performance based on financial and operational indicators. These models help bank managers and regulators better understand efficiency drivers and improve decision-making [25]. Non-parametric ML algorithms are particularly suitable, as they can flexibly model non-linear and non-parametric relationships between bank efficiency and financial variables [25].

In Vietnam, the application of ML in banking and finance is gradually expanding. Previous studies applied ML to early warning systems for debt reclassification in Vietnamese commercial banks [26]. Other research combined DEA and ML to predict the performance of Vietnamese micro, small, and medium enterprises, showing higher accuracy than traditional econometric methods [27]. ML-based credit scoring models for individual customers at Vietnamese banks have also attracted increasing research attention [28].

2.4. Hyperparameter Optimization

Most ML algorithms rely on a set of hyperparameters whose values need to be carefully chosen, as they can substantially affect model performance [29], [30]. Hyperparameter optimization (HPO) refers to the process of systematically searching for an appropriate hyperparameter configuration for a given ML model [30], [31]. Existing HPO techniques range from relatively simple approaches, such as grid search and random search, to more advanced methods including Bayesian optimization, evolutionary algorithms, Hyperband, and Racing [29], [30]. The main purpose of HPO is to reduce the reliance on time-consuming and often irreproducible manual trial-and-error procedures when identifying well-performing hyperparameter settings [30]. Therefore, selecting a suitable HPO technique plays an important role in determining the final performance of ML models [29].

2.5. Interpreting Machine Learning Models with SHAP

In highly regulated fields such as banking and finance, understanding how ML models generate their predictions is essential to ensure transparency and accountability [3]. Although many black-box models deliver strong predictive performance, they often provide limited explanation of the factors driving their results. To address this issue, SHAP has been introduced as a general framework for interpreting the outputs of ML models [3], [32], [33].

SHAP assigns an importance value to each feature for a specific prediction based on Shapley values derived from cooperative game theory [34], [35]. A feature’s Shapley value is calculated as its average marginal contribution across all possible feature combinations, which ensures a fair distribution of the model output among explanatory variables [36], [37]. This framework satisfies several desirable properties, including efficiency, symmetry, and linearity [38], and can be applied to different ML models in a consistent manner [35].

For tree-based algorithms, variants such as TreeSHAP reduce computational complexity while retaining exact Shapley value estimation. Features with larger absolute Shapley values are therefore regarded as more influential predictors [36]. In addition, SHAP summary plots help rank variables by importance and illustrate how changes in explanatory variables affect predicted outcomes, thereby revealing non-linear relationships and supporting economic interpretation of model results [38], [39].

Several studies have combined DEA and machine learning methods to examine bank efficiency. These contributions are valuable, but some gaps remain. First, many studies focus mainly on predictive accuracy. Model performance is often evaluated using random splits or standard cross-validation. In financial data, where observations follow a time order, this may not reflect real forecasting conditions; Second, interpretation is usually limited. Feature importance is reported, but deeper analysis of non-linear effects, threshold behavior, or interaction patterns is less common. As a result, the economic meaning of predictions is not always fully explored; Third, efficiency measurement and prediction are often treated as separate steps. A clear analytical sequence that links efficiency estimation, time-aware validation,

ISSN 2723-6471

robustness checks, and structured interpretation is still limited in the literature. This study addresses these points by developing an integrated framework that connects these elements within a single design.

3. Methodology

3.1. Data sources and collection

The sample includes 27 Vietnamese commercial banks observed from 2016 to 2024. Data were obtained from the FiinproX database and the banks' audited annual financial statements. This study does not utilize interim, projected, or forecast figures. All observations are based on officially reported year-end financial data to ensure consistency and comparability across years. The period 2016–2024 reflects a significant transformation in the Vietnamese banking system. During this time, banks engaged in addressing large-scale non-performing loans, restructuring the system, gradually adopting Basel standards, responding to the COVID-19 pandemic, and accelerating digital transformation. These structural and regulatory changes may contribute to the variability in efficiency dispersion between years, particularly in the later period of the study sample. Because efficiency scores are calculated from audited annual data, they reflect actual efficiency rather than projected results. Future research could expand the sample to 2025 when audited financial statements for that year are published.

To develop a reliable bank efficiency forecasting model, it is essential to carefully select input variables that accurately capture the key drivers of banks' operational and financial performance. Variable Selection for the DEA-DDF Efficiency Estimation Model (table 1):

Table 1. Variable Selection for the DEA-DDF Efficiency Estimation Model

Category	Variable
Approach	Financial intermediation. Banks are modeled as financial intermediaries transforming deposits into loans and income; appropriate for Vietnamese banks where 70–80% of funding comes from customer deposits [10].
Inputs	Customer deposits
	Personnel & management expenses
	Interest and similar expenses
	Equity
Desirable outputs	Customer loans
	Non-interest income
	Interest and similar income
Undesirable output	Loan loss provisions (LLP)

Under the intermediation approach, banks use funding sources to generate earning assets. Customer deposits and interest expenses represent major funding inputs. Personnel and management expenses reflect operational input.

Equity is included as an input because it provides internal capital that supports lending and absorbs risk. Here, equity is treated as part of the production process. It is not the same as the capital adequacy ratios used later in the machine learning stage. Equity reflects the level of internal funding, while capital adequacy measures regulatory capital relative to risk-weighted assets. Selection of Predictive Features (table 2):

Table 2. Definition of variables used in the ML efficiency forecasting models

Variable	Definition	Expected
ROA	Return on asset; Measures overall asset utilization efficiency and reflects managerial and operational performance [7]	Positive
NIM	Net-interest-margin; Captures profitability from interest-based activities and pricing efficiency [1]	Positive
CIR	Cost-to-income ratio; lower values indicate better cost control and higher efficiency [25]	Negative
NII	Non-interest income; reflects revenue diversification beyond traditional lending [4]	Positive
NPL	Non-performing loan ratio; higher credit risk increases provisioning costs and reduces efficiency [24]	Negative
CAR	Capital adequacy ratio; enhances stability but may constrain efficiency if excessively high [2]	Ambiguous
Leverage	Reliance on borrowed funds; may improve efficiency but increases financial risk [4]	Ambiguous
LDR	Loan-to-deposit ratio; indicates liquidity utilization and intermediation intensity [1]	Ambiguous

ln_assets	Natural logarithm of total assets; Controls for scale effects and data distribution [1]	Ambiguous
GDP	GDP captures overall economic activity and macroeconomic conditions affecting banking operations [40]	Positive
INF	Inflation rate; affects interest rates, costs, and asset values [40]	Mixed
CAR × ROA	Captures moderation effect of capital adequacy on profitability–efficiency relationship [39]	Conditional

3.2. DEA – DDF method for efficiency estimation

This subsection represents the first pillar of the research framework: measuring bank efficiency before moving to the forecasting stage. Bank efficiency is estimated using Data Envelopment Analysis with a Directional Distance Function (DEA–DDF). The financial intermediation approach is adopted. Banks are viewed as institutions that transform inputs such as deposits, equity, and operating expenses into outputs such as loans and income. In addition to desirable outputs, the model incorporates undesirable outputs related to credit risk, especially loan loss provisions. This reflects the fact that bank performance depends not only on income generation but also on risk control. The directional distance function is defined as follows:

$$\vec{D}(x, y, b; g_y, g_b) = sup\,\{\beta: (x, y + \beta g_{y,} b - \beta g_b) \in T\}. \quad (1)$$

where $x_i \in R_+^M$, $y_i \in R_+^S$, and $b_i \in R_+^Q$ denote the input, desirable output,and the undesirable output vectors for bank $i = 1 , \ldots , N$. The direction vector is defined as $g = (g_y, -g_b)$, where $g_y \in R_+^S$ and $g_b \in R_+^Q$. The scalar $\beta \geq 0$ measures the maximum feasible expansion of desirable outputs and contraction of undesirable outputs along the chosen direction. The production possibility set $T \subseteq R_+^{M+S+Q}$ is constructed from the observed sample under the variable returns to scale (VRS) assumption. The efficiency score (θ) is computed as $1 - \beta$, with values ranging from 0 to 1.

The direction vector is specified to increase desirable outputs and reduce undesirable outputs simultaneously. This aligns with the practical objective of commercial banks, which seek to expand income while managing credit risk. A proportional adjustment is applied to allow consistent comparison across banks of different sizes. Alternative directional choices are possible, such as input reduction only. However, in a risk-adjusted setting, joint adjustment of desirable and undesirable outputs better reflects banking operations. The focus of this study is on relative efficiency patterns rather than exact efficiency levels.

To improve reliability, a bootstrap procedure following Simar and Wilson is applied to the estimated efficiency scores. Let θ_i denote the original efficiency score and $\overline{\theta_i^*}$ denote the mean of the bootstrap estimates. The estimated bias is computed as:

$$bias_i = \overline{\theta_i^*} - \theta_i \quad (2)$$

The bias-corrected efficiency score is then:

$$\theta_i^{BC} = \theta_i - bias_i \quad (3)$$

This adjustment addresses the upward bias commonly observed in conventional DEA estimators. For banks located on the frontier $(\theta_i = 1)$, bootstrap resampling may produce slightly lower mean estimates, leading to small downward corrections. These adjustments reflect sampling variability rather than structural inefficiency. Super-efficiency is not applied in this study. Because the efficiency scores are later used as dependent variables in the forecasting stage (Pillar 2), they are normalized to ensure compatibility with supervised learning algorithms.

The DEA–DDF model is implemented in Python using the Pyomo optimization framework. Linear programming problems are solved with the open-source CBC solver via Pyomo's SolverFactory interface. Optimization is conducted under the VRS assumption. Feasibility and optimality tolerances are set to 10^{-6}. All decision variables are restricted to non-negative values. Computations are performed in Google Colab. The model specification and solver configuration are kept consistent across years to ensure reproducibility.

3.3. Machine Learning models used for forecasting

Owing to the complex and non-linear nature of financial data, ML models are increasingly used to predict banking efficiency [20], [21]. Compared with traditional econometric models, ML techniques better capture non-linear relationships and deliver higher predictive accuracy [22]. This study therefore adopts tree-based ensemble models for panel data, which have demonstrated strong performance in financial forecasting tasks [25].

 ISSN 2723-6471

This subsection corresponds to the second pillar of the research framework: forecasting the efficiency scores obtained in Pillar 1. After bank efficiency is estimated using the DEA–DDF, the resulting scores are treated as dependent variables in supervised learning models. They are not directly observed values. For this reason, they may contain some degree of measurement error. This can result from sample variation, model assumptions, or the choice of direction vector.

When these estimated scores are later used as dependent variables in the machine learning stage, it is important to clarify the objective of the analysis. The purpose here is prediction, not causal inference. The study does not attempt to estimate structural relationships or draw conclusions about economic causality. Instead, it examines how well observable financial indicators can forecast relative efficiency patterns across banks.

To improve reliability, a bootstrap procedure is applied to obtain bias-corrected efficiency estimates and confidence intervals. This helps reduce instability in the efficiency scores. Even so, the estimated values should be interpreted as relative indicators of performance rather than exact measures of true efficiency. The ML models considered include:

(1) Linear Models: Classical linear regression models, while simple and interpretable, often lack the capacity to capture the intricate complexities of banking efficiency. Variants such as Ridge and Lasso Regression can improve generalization through regularization terms but are still limited by linearity assumptions.

(2) Tree-based models are better suited for capturing non-linear effects. Decision Trees are flexible but prone to overfitting. Random Forest reduces this risk by averaging many trees. Gradient Boosting further improves accuracy by building trees sequentially. Among boosting methods, XGBoost is widely used due to its efficiency and strong predictive performance in banking studies [25].

(3) Support Vector Machines perform well in high-dimensional settings but require careful kernel and parameter selection.

(4) Neural networks can model complex non-linear relationships and time dependence. However, they often require larger datasets and are harder to interpret.

The selection of these models is based on their ability to address the characteristics of banking data, such as non-linearity and time-varying dynamics. For this study, the interpretation phase focused on the best-performing model to avoid duplication.

3.4. Time – series Validation

This subsection represents the third pillar of the research framework: evaluating model performance in a realistic time-dependent setting. Financial data are observed over time, and this simple fact affects how models should be evaluated. Standard methods such as k-fold cross-validation assume that observations are independent. In time-series settings, this assumption is rarely realistic. When k-fold is applied to time-ordered data, information from future periods may leak into the training process. This can make model performance look stronger than it actually is [41], [43]. Earlier studies have also pointed out that k-fold validation is only appropriate under rather restrictive conditions in time-dependent contexts [44].

To avoid this issue, time-series cross-validation is adopted using a rolling-origin structure through TimeSeriesSplit [40], [42]. Models are trained on earlier years and tested on later ones. The training window expands gradually, and the test period always follows the training period. This preserves the temporal order of the data and ensures that predictions rely only on information that would have been available at that time.

Time-series validation is implemented using TimeSeriesSplit with five splits. An expanding-window approach is applied. In each fold, the model is trained on all available past years and tested on the next year. For example, the first fold trains on the earliest year and tests on the following year. The second fold trains on the first two years and tests on the third year. The training window expands over time. The test window advances one year at a time. No rolling or fixed-size window is used. This design reflects real forecasting practice and avoids information leakage. This approach often produces lower performance metrics compared to standard cross-validation. However, it provides a more realistic assessment of forecasting performance in practice [43]. For this reason, it is preferred in the present study.

3.5. Hyperparameter optimization techniques

Model performance depends not only on the data but also on how models are configured. Algorithms such as XGBoost include several hyperparameters that influence tree depth, learning rate, regularization, and sampling behavior. Poor parameter choices may lead to overfitting or weak generalization [29], [30]. Therefore, hyperparameter tuning is treated as a necessary part of the forecasting procedure [30], [31].

In this study, hyperparameters are selected using a two-stage search procedure. RandomizedSearchCV is first applied to explore the parameter space efficiently. GridSearchCV is then conducted on a reduced grid for the best-performing model to refine the final configuration. All tuning procedures are carried out within the TimeSeriesSplit framework described in Section 3.4. In each of the five splits, only past observations are used for training and parameter selection, while the subsequent year is used for validation.

Although Bayesian optimization has been suggested in the literature as an efficient alternative in complex search settings [29], [31], it is not implemented here. The current approach prioritizes transparency, stability, and reproducibility under strict time-dependent validation. Beyond parameter tuning, robustness is also examined. Model performance is evaluated across different temporal splits to ensure that results are not driven by a specific training period. Near-optimal parameter configurations produce similar performance outcomes, suggesting that the results are reasonably stable. After establishing predictive performance and robustness, the next step is to understand how the model arrives at its predictions. This leads to the fourth pillar of the framework.

3.6. Model Interpretability using SHAP

The fourth pillar focuses on interpretation. In highly regulated sectors such as banking, model transparency is essential for trust and accountability [3]. While complex machine learning models can improve prediction accuracy, their internal mechanisms are not always easy to understand. Without interpretation, performance metrics alone provide limited insight. To address this issue, SHAP is used as an interpretation framework [3], [32], [33]. SHAP is grounded in cooperative game theory and applies Shapley values to quantify the contribution of each feature to an individual prediction [34], [35]. A SHAP value indicates how much a specific variable shifts the predicted efficiency away from a baseline level [45], [46]. Positive values increase the prediction, while negative values decrease it [32], [33].

Because the forecasting stage relies primarily on XGBoost, TreeExplainer is employed for computational efficiency and suitability for tree-based models [32], [33]. SHAP allows interpretation at multiple levels. Global importance rankings identify key drivers across the sample. Local explanations clarify how individual variables affect a specific bank-year prediction. Dependence plots reveal non-linear patterns, and interaction values capture joint effects between variables—an important feature in financial settings where factors rarely operate in isolation [39], [47]. By integrating interpretation into the analytical sequence, the study moves from measuring efficiency, to predicting it, to validating performance, and finally to explaining the results in economic terms. This completes the four-pillar framework and ensures that predictive accuracy is complemented by meaningful economic insight.

3.7.Research Framework

The overall research framework is presented in figure 1. The procedure consists of four main stages. First, efficiency scores are estimated using Data Envelopment Analysis with Directional Distance Function. Bootstrap bias correction is applied to improve reliability. Second, the corrected efficiency scores are used as target variables in machine learning models. Feature construction and time-series validation are performed to ensure realistic forecasting.Third, hyperparameter optimization and model evaluation are conducted to identify the best-performing model. Finally, Shapley value analysis is applied to interpret the prediction results and identify key efficiency drivers.

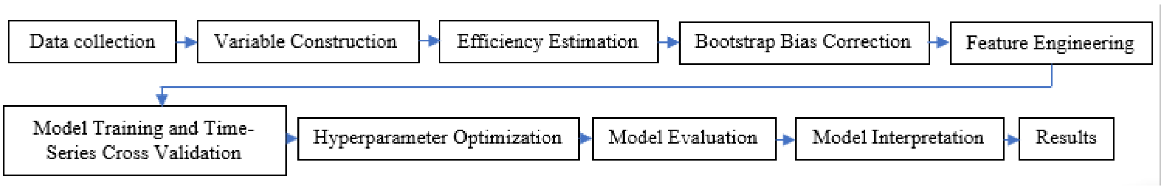

Figure 1. Research framework for efficiency estimation, forecasting, validation, and model interpretation

3.8. Pseudocode of the Proposed Framework

The proposed framework follows a sequential procedure. First, bank efficiency scores are estimated using Data Envelopment Analysis with Directional Distance Function. Bootstrap bias correction is then applied to improve the reliability of the efficiency scores. Next, the bias-corrected efficiency scores are combined with financial indicators to construct the predictive dataset. The data are split using time-series cross-validation to preserve the chronological order.

Machine learning models are trained, and hyperparameter optimization is performed to improve predictive performance. Model accuracy is evaluated using root mean squared error, mean absolute error, and coefficient of determination. The best-performing model is then selected. Finally, Shapley value analysis is applied to interpret the prediction results and identify the key determinants of efficiency. The framework produces predicted efficiency scores and their corresponding interpretation results.

4. Results and Discussion

4.1. Bank Efficiency analysis using DEA - DDF

The research results, summarized in table 3, show that the average efficiency of 27 Vietnamese banks remains at a very high level, ranging from 0.9942 to 0.9987. In the period 2016-2018, although the average efficiency was 0.994, the standard deviation was relatively large, especially in 2017, reaching 0.0198, reflecting significant differences among banks with a minimum value of only 0.902. From 2019 to 2021, efficiency gradually improved and dispersion decreased. However, from 2022 to 2024, the standard deviation increased again, showing signs of renewed differentiation, with some banks exhibiting lower efficiency. This result implies that, although the Vietnamese banking system maintains high and stable operational efficiency, new fluctuations in 2024 warrant attention, potentially stemming from macroeconomic pressures, capital increase requirements, or increased credit risks.

Table 3. DEA – DDF Efficiency Scores

Year	Count	Mean	Std	Min	Max
2016	27	0.9949	0.0147	0.9400	1
2017	27	0.9944	0.0198	0.9020	1
2018	27	0.9942	0.0151	0.9402	1
2019	27	0.9955	0.0171	0.9186	1
2020	27	0.9967	0.0169	0.9122	1
2021	27	0.9968	0.0088	0.9647	1
2022	27	0.9979	0.0077	0.9702	1
2023	27	0.9987	0.0048	0.9812	1
2024	27	0.9970	0.0121	0.9379	1

To confirm the reliability of the DEA–DDF efficiency scores, this study used Simar & Wilson's Bootstrap DEA analysis. As presented in table 4, the Bootstrap application results show that the bias-corrected efficiency scores are slightly lower than the original results, implying that DEA tends to be more optimistic than actual efficiency. However, the difference between the original and bias-corrected efficiency is generally not large, and the 95% confidence interval is narrow, indicating the robustness and stability of the model. This strengthens the reliability of combining DEA–DDF with Bootstrap in measuring bank efficiency.

Table 4. DEA – DDF Efficiency Scores

Bank name	DEA – DDF	Mean Bootstrap Efficiency	Bias-Corrected Efficiency	95% CI Lower	95% CI Upper
Vietnam Bank for Agriculture and Rural Development	1	0.9978	0.9978	0.9652	1
An Binh Commercial Joint Stock Bank	0.9761	0.9979	0.9978	0.9673	1
Bao Viet Joint Stock Commercial Bank	0.9993	0.9978	0.9977	0.9652	1
Bac A Commercial Joint Stock Bank	0.9972	0.9977	0.9977	0.9638	1
Vietnam Joint Stock Commercial Bank for Industry and Trade	1	0.9979	0.9979	0.9677	1
Vietnam Maritime Commercial Joint Stock Bank	0.9874	0.9982	0.9982	0.96691	1
Kien Long Commercial Joint Stock Bank	1	0.9977	0.9977	0.9638	1

ISSN 2723-6471

Vietnam Technological and Commercial Joint Stock Bank	1	0.9979	0.9979	0.9676	1
Nam A Commercial Joint Stock Bank	1	0.9978	0.9978	0.9647	1
Joint Stock Commercial Bank for Foreign Trade of Vietnam	0.9931	0.9977	0.9978	0.9635	1
Ho Chi Minh City Development Joint Stock Commercial Bank	1	0.9980	0.9980	0.9677	1
Orient Commercial Joint Stock Bank	1	0.9978	0.9978	0.9643	1
Military Commercial Joint Stock Bank	1	0.9978	0.9978	0.9643	1
National Citizen Bank	0.9965	0.9979	0.9979	0.9645	1
Vietnam International Commercial Joint Stock Bank	1	0.9978	0.9978	0.9673	1
Saigon Bank for Industry and Trade	1	0.9978	0.9978	0.9657	1
Saigon Thuong Tin Commercial Joint Stock Bank	0.9784	0.9979	0.9979	0.9661	1
Saigon-Hanoi Commercial Joint Stock Bank	1	0.9977	0.9977	0.9650	1
Tien Phong Commercial Joint Stock Bank	1	0.9977	0.9977	0.9641	1
Vietnam Joint Stock Commercial Bank for Industry and Trade	0.9979	0.9979	0.9979	0.9660	1
Vietnam Prosperity Joint Stock Commercial Bank	1	0.9979	0.9980	0.9685	1
Viet A Commercial Joint Stock Bank	1	0.9978	0.9978	0.9647	1
Vietnam Export Import Commercial Joint Stock Bank	0.9719	0.9979	0.9979	0.9643	1
Petrolimex Commercial Joint Stock Bank	1	0.9979	0.9979	0.9654	1
Asia Commercial Joint Stock Bank	1	0.9977	0.99772	0.9647	1
Southeast Asia Commercial Joint Stock Bank	0.9993	0.9978	0.9979	0.9643	1
Joint Stock Commercial Bank for Investment and Development of Vietnam	1	0.9978	0.9979	0.9643	1

The DEA–DDF scores are high and clustered near 1. This ceiling effect limits variation in the dependent variable and makes forecasting small changes more difficult. Even so, differences across banks and years remain, especially in periods with higher dispersion. The forecasting stage therefore focuses on relative movements rather than large efficiency gaps.

4.2. Bank Efficiency Forecasting Results Using Machine Learning

This section details the findings from the construction, evaluation, and interpretation of the bank operational efficiency forecasting model. Results are presented sequentially, covering the comparison of ML model performance, robustness assessment using time-series cross-validation, hyperparameter optimization, and feature importance and interpretation using SHAP.

4.2.1. Comparison of Forecasting Model Performance.

To identify the optimal ML model for bank efficiency forecasting, we compared the performance of five different models on the test set with engineered features. Evaluation metrics included Root Mean Squared Error, Mean Absolute Error, and Coefficient of Determination (R^2). Table 4 summarizes the performance evaluation results.

Table 5. Performance Evaluation Results of Models on the Test Set

Model	RMSE	MAE	R^2
XGBoost	0.060107	0.047900	0.352567
RandomForest	0.064299	0.051457	0.259098
LightGBM	0.065785	0.050100	0.224459
CatBoost	0.071048	0.058242	0.095404
NeuralNet	0.207100	0.149495	-6.686176

As shown in table 5, the XGBoost model demonstrated superior performance compared to the other models, achieving the lowest RMSE of 0.060107, the lowest MAE of 0.047900, and the highest $R^2$of 0.352567. This confirms the effectiveness of Gradient Boosting algorithms in handling tabular data and capturing complex, non-linear relationships in financial efficiency forecasting, consistent with previous research [25]. In contrast, the Neural Network model showed very poor performance with a significantly negative R^2 , suggesting it is either unsuitable for this dataset or requires a much more complex architecture and optimization to achieve acceptable performance.

Although several models were evaluated, XGBoost achieved the most stable and competitive performance across validation settings. Therefore, subsequent interpretation focuses on XGBoost.

4.2.2. Time – Series Validation Results

To assess robustness and generalization in a time-dependent setting, we applied TimeSeriesSplit cross-validation [40], [42]. This method preserves temporal order and avoids information leakage [43], [44]. For XGBoost, the average results were: RMSE = 0.084607, MAE = 0.063180, and R^2 = 0.004787.

Model performance declined compared to the static test evaluation. R^2 dropped from 0.352567 to 0.004787, indicating limited explanatory power under strict time-series validation. The efficiency scores are bounded between 0 and 1 and clustered near the frontier, which leaves little variation to explain. The objective here is not to predict large efficiency gaps but to track relative changes over time. Even modest predictive power may still provide useful signals. This result also reflects the difficulty of forecasting financial time-series data, where relationships change and volatility is high. It suggests the need for more dynamic, time-dependent features.

4.2.3. Hyperparameter Optimization Techniques

To ensure the XGBoost model operates at stable performance and avoids overfitting, we applied a two-stage tuning procedure. RandomizedSearchCV was first used to explore the parameter space, and GridSearchCV was then conducted on a reduced grid to refine the final configuration [30], [31] The optimal hyperparameter set found for the XGBoost model is: colsample_bytree: 0.8, learning_rate: 0.07, max_depth: 3, n_estimators: 600, subsample: 0.8. Tuning these hyperparameters helps balance model complexity and generalization ability [29].

4.2.4. Feature Importance Analysis

To better understand which factors contribute most to the XGBoost model's predictions, we calculated feature importance using the Permutation Importance method. This method assesses how much the model's performance decreases when the values of a specific feature are randomly shuffled. Based on Permutation Importance, the most important features are ranked in descending order as follows: ROA, NIM, ln_assets, CIR, CARxROA (figure 2).

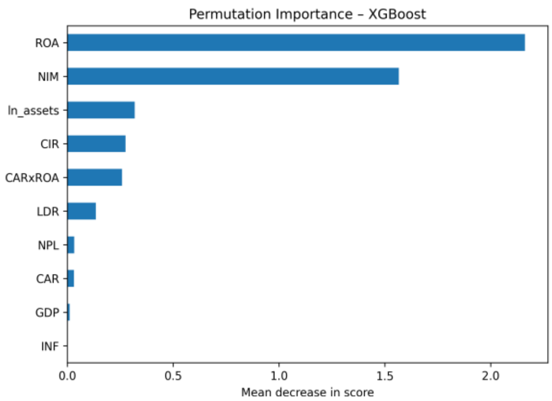

Figure 2. Permutation importance of input features in the XGBoost model

This ranking provides an initial insight into the variables that have the greatest influence on bank efficiency forecasts according to the XGBoost model, with profitability and size indicators standing out.

4.2.5. Interpreting Models with SHAP

To enhance transparency in the prediction mechanism of the XGBoost model and to clarify how individual input features influence DEA–DDF efficiency scores, this study applies the SHAP framework [3], [32], [33]. SHAP decomposes each prediction into additive contributions from individual features, enabling both local (instance-level) and global (model-level) interpretability. Given the tree-based nature of XGBoost, TreeExplainer is employed to efficiently compute SHAP values.

SHAP force plots illustrate how individual features shift predicted efficiency scores above or below the baseline, defined as the sample mean prediction, with red and blue segments indicating negative and positive contributions, respectively. Low – efficiency cases are primarily driven by high CIR and low profitability (ROA), which pull predictions below the baseline, whereas high-efficiency cases are dominated by strong positive contributions from ROA and NIM alongside low CIR. Observations with large prediction errors often exhibit conflicting signals, such as high profitability combined with adverse interaction effects, suggesting that efficiency outcomes may reflect complex interactions or unobserved factors. In contrast, randomly selected cases tend to display a balanced mix of positive and negative contributions, consistent with average banking conditions.

Figure 3 presents the SHAP summary plot provides a global overview of feature importance and the direction of their effects across all observations [45]. Features are ranked by decreasing importance, with ROA and NIM emerging as the most influential determinants of predicted efficiency, followed by bank size (ln_assets), CIR, and the interaction term CAR×ROA. Higher ROA values are associated with positive SHAP values, consistently pushing efficiency predictions upward, confirming profitability as a core driver of operational efficiency. NIM exhibits a largely monotonic positive effect, though signs of saturation appear at higher levels, indicating diminishing marginal gains. The effect of ln_assets is relatively neutral or mildly positive, suggesting that bank size alone does not guarantee higher efficiency. CIR displays a strong negative association, with higher cost burdens reducing predicted efficiency. The interaction term CAR×ROA contributes conditionally, implying that the combined effect of capitalization and profitability may weaken efficiency under certain configurations.

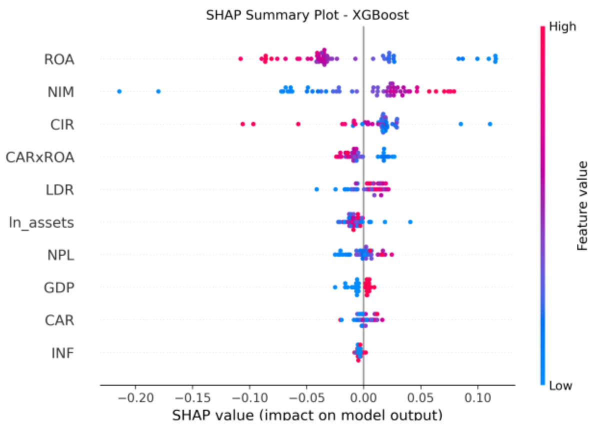

Figure 3. SHAP summary plot showing global feature importance and directional effects

Figure 4 presents the SHAP dependence plots provide detailed insights into how individual features influence model predictions and reveal non-linear patterns and thresholds [46]. The dependence plot for NIM confirms its positive relationship with efficiency, while also indicating diminishing returns at very high levels. CIR exhibits a clear negative relationship, with a pronounced decline threshold around 0.55, beyond which efficiency deteriorates more sharply. The plot for ln_assets confirms a near-neutral effect across most of its range, suggesting that economies of scale may be offset by managerial complexity. For the interaction term CAR×ROA, the plots indicate a conditionally negative effect when the balance between capitalization and profitability becomes misaligned, highlighting the importance of interaction effects in explaining banking efficiency [39]. Overall, SHAP analysis not only validates the importance of core financial indicators but also uncovers complex non-linear dynamics and interactions, providing valuable insights for bank management and regulatory assessment [36], [37].

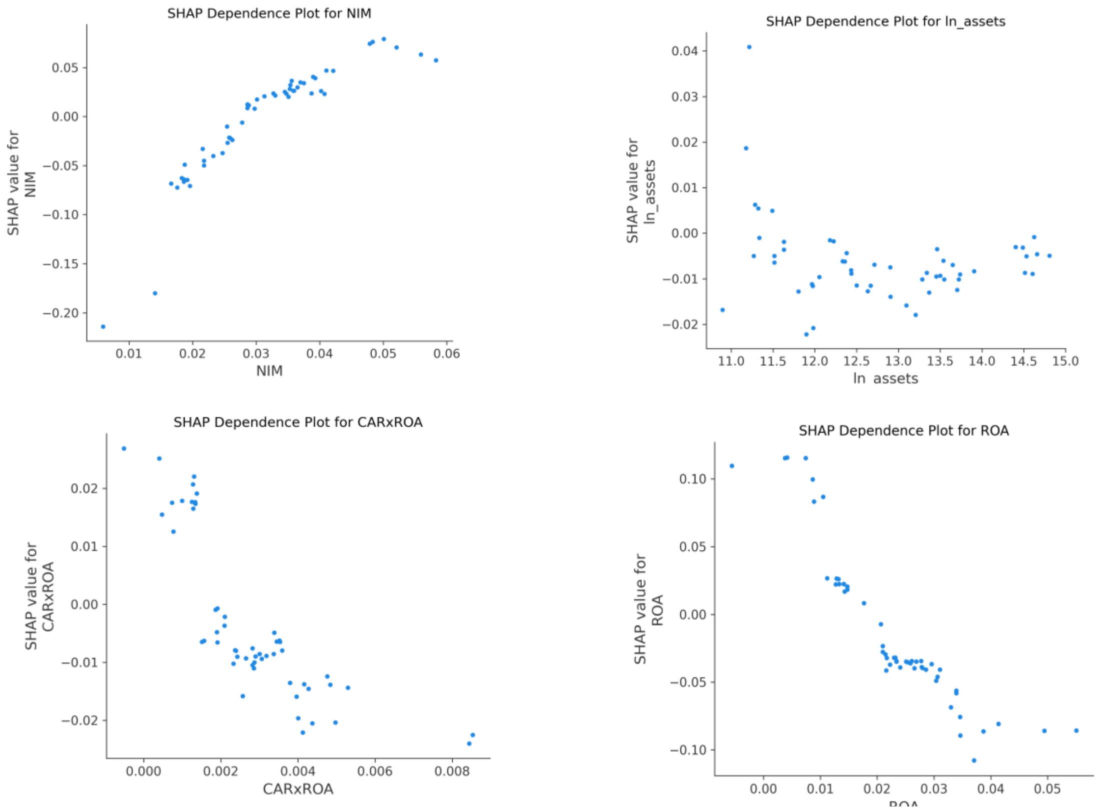

Figure 4. SHAP dependence plots for NIM, ROA, ln_assets, CARxROA

The SHAP findings are compared with the expected signs reported in table 2 and the supporting literature. The positive contribution of ROA is consistent with prior evidence that higher asset utilization reflects stronger managerial and operational performance [7]. Similarly, the positive impact of NIM aligns with earlier findings that interest-based profitability supports efficiency [1]. The negative effect of CIR is also in line with existing studies showing that weaker cost control reduces efficiency [25]. These results suggest that the empirical SHAP patterns broadly support the theoretical expectations. For variables with ambiguous or conditional expectations, the SHAP results provide additional nuance. The near-neutral effect of ln_assets is consistent with the mixed evidence on scale effects [1]. The interaction term CAR×ROA exhibits a conditional pattern, which is aligned with the moderating role of capital adequacy discussed in prior research [39].

For other variables such as NII [4], NPL [24], GDP and INF [40], the SHAP contributions are comparatively smaller in this model version. Therefore, their directional effects are not over-interpreted. In summary, the SHAP analysis largely supports the expected signs while also revealing non-linear and interaction effects that extend earlier linear findings.

4.2.6. Results of Robustness Checks

Table 6 compares the performance of the XGBoost model under the baseline data split and various robustness check scenarios. To assess the stability and generalizability of DEA–DDF efficiency forecasts, robustness checks are conducted by varying temporal data partitions and feature selection strategies, which is critical in dynamic financial environments subject to structural change [41], [43]. The baseline model is trained on data from 2016–2021 and evaluated on 2022–2024 (year_cut = 2022), while alternative splits at year_cut = 2021 and year_cut = 2020 progressively extend the forecast horizon and reduce the training window, creating more stringent time-series prediction settings [40]. In addition, feature subset sensitivity is examined by excluding the three least important variables (GDP, INF, and CAR) based on permutation importance rankings. The model is re-estimated using the reduced feature set under the year_cut = 2021 configuration to evaluate whether a more parsimonious specification enhances predictive accuracy and robustness [21].

Table 6. Comparative Performance of XGBoost under Robustness Checks

Model	RMSE	MAE	R^2	Split_type
XGBoost	0.060107	0.047900	0.352567	year_cut=2022
XGBoost	0.067495	0.054180	0.047433	year_cut=2021
XGBoost	0.063830	0.051518	0.148070	year_cut=2021
XGBoost	0.077427	0.062821	-0.147235	year_cut=2020

Note: RMSE and MAE measure prediction errors, while R^2 captures explanatory power. Reduced-feature models exclude GDP, INF, and CAR.

The robustness results indicate strong sensitivity to temporal data partitioning. As the training window shortens and the forecasting horizon lengthens, predictive performance deteriorates markedly, with R^2 falling from 0.35 in the baseline specification to near zero and turning negative under the most restrictive split. This pattern highlights the challenge of capturing evolving data-generating processes in financial time series [43]. By contrast, excluding the least informative variables leads to modest performance gains under the same temporal split, suggesting that a more parsimonious feature set can reduce noise and limit overfitting [21]. Overall, the findings imply that while core financial indicators remain central, stable long-horizon forecasting requires richer time-dependent features, such as lagged variables, rolling measures, or regime-sensitive structures, in line with the proposed model enhancement pathway.

5. Conclusion

An integrated framework combining DEA–DDF efficiency measurement, advanced machine learning techniques, and SHAP-based interpretability is employed to analyze and forecast the operational efficiency of 27 Vietnamese commercial banks during the period 2016–2024. By jointly addressing efficiency estimation, prediction, and interpretability, the framework responds to both methodological challenges and practical demands in banking efficiency analysis.

Empirical evidence shows that XGBoost achieves the highest predictive accuracy under static evaluation settings, supporting the effectiveness of gradient boosting methods in capturing non-linear efficiency patterns in banking data. SHAP results indicate that profitability indicators, particularly ROA and NIM, are the most influential drivers of efficiency, while cost efficiency exerts a significant impact through a threshold effect in the Cost-to-Income Ratio. Bank size appears largely neutral, whereas the interaction between capital adequacy and profitability suggests a trade-off between financial resilience and operational efficiency.

Robustness analysis based on time-series cross-validation reveals pronounced sensitivity to temporal data partitioning, with predictive performance declining as the forecasting horizon extends. This finding highlights the limitations of static feature representations in dynamic financial environments and emphasizes the importance of time-aware modeling strategies.

Taken together, the findings contribute to the banking efficiency literature by providing an interpretable and empirically grounded forecasting framework with practical implications for bank managers and policymakers. Future research may focus on dynamic feature construction, advanced optimization methods, and dedicated time-series models to improve forecasting stability and support more effective supervisory and strategic decision-making.

6. Declarations

6.1. Author Contributions

Conceptualization: C.D.L; Methodology: C.D.L; Software: C.D.L; Validation: C.D.L; Formal Analysis: C.D.L; Investigation: C.D.L; Resources: C.D.L; Data Curation: C.D.L; Writing Original Draft Preparation: C.D.L; Writing Review and Editing: C.D.L; Visualization: C.D.L; All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Ö. O. Akdeniz, H. A. Abdou, A. Hayek, J. C. Nwachukwu, A. A. Elamer, and C. Pyke, "Technical efficiency in banks: a review of methods, recent innovations and future research agenda," *Review of Managerial Science*, vol. 18, no. 11. Springer Science+Business Media, p. 3395, Dec. 21, 2023. doi: 10.1007/s11846-023-00707-z.

[2] K. D. Dutta and M. Saha, "Do competition and efficiency lead to bank stability? Evidence from Bangladesh," *Future Business Journal,* vol. 7, no. 1, Jan. 2021, doi: 10.1186/s43093-020-00047-4.

[3] S. Lundberg and S. Lee, "A Unified Approach to Interpreting Model Predictions," *arXiv (Cornell University)*, May 2017, doi: 10.48550/arxiv.1705.07874.

[4] M. Izzeldin, E. Mamatzakis, A. Murphy, V. Pappas, and E. G. Tsionas, "An innovative Bayesian multiple indicator-multiple cause analysis of bank productivity," *Review of Quantitative Finance and Accounting*, Oct. 2025, doi: 10.1007/s11156-025-01445-x.

[5] M. Koetter, "The Stability of Efficiency Rankings When Risk-Preferences and Objectives are Different," *SSRN Electronic Journal,* Jan. 2006, doi: 10.2139/ssrn.2793981.

[6] J. P. Hughes and L. J. Mester, "Measuring the Performance of Banks: Theory, Practice, Evidence, and Some Policy Implications," *SSRN Electronic Journal,* Jan. 2013, doi: 10.2139/ssrn.2306003.

[7] J. P. Hughes and L. J. Mester, "Efficiency in Banking: Theory, Practice, and Evidence," *Federal Reserve Bank of Philadelphia,* Jan. 2008. doi: 10.21799/frbp.wp.2008.01.

[8] X. Zhou, Z. Xu, J. Chai, L. Yao, S. Wang, and B. Lev, "Efficiency evaluation for banking systems under uncertainty: A multi-period three-stage DEA model," *Omega*, vol. 85, p. 68, May 2018, doi: 10.1016/j.omega.2018.05.012.

[9] A. P. Balcerzak et al., "Non-Parametric Approach to Measuring the Efficiency of Banking Sectors in European Union Countries," *Acta Polytechnica Hungarica,* vol. 14, no. 7, Feb. 2018, doi: 10.12700/aph.14.7.2017.7.4.

[10] J. Tanwar, H. Seth, A. K. Vaish, and N. V. M. Rao, "Revisiting the Efficiency of Indian Banking Sector: An Analysis of Comparative Models Through Data Envelopment Analysis," *Indian Journal of Finance and Banking*, vol. 4, no. 1, p. 92, May 2020, doi: 10.46281/ijfb.v4i1.585.

[11] R. Färe, S. Grosskopf, and D. Margaritis, "Directional Distance Functions Revisited: Selective Overview and Update," *Data Envelopment Analysis Journal*, vol. 1, no. 2, p. 57, Jan. 2015, doi: 10.1561/103.00000006.

[12] R. Färe and S. Grosskopf, "Theory and Application of Directional Distance Functions," *Journal of Productivity Analysis*, vol. 13, no. 2, p. 93, Mar. 2000, doi: 10.1023/a:1007844628920.

[13] C. Daraio, L. Simar, and P. W. Wilson, "Fast and efficient computation of directional distance estimators," *Annals of Operations Research*, vol. 288, no. 2, p. 805, Feb. 2019, doi: 10.1007/s10479-019-03163-9.

[14] B. K. Sahoo, M. Mehdiloozad, and K. Tone, "Cost, revenue and profit efficiency measurement in DEA: A directional distance function approach," *European Journal of Operational Research*, vol. 237, no. 3, p. 921, Feb. 2014, doi: 10.1016/j.ejor.2014.02.017.

[15] C. Yang, "An enhanced DEA model for decomposition of technical efficiency in banking," *Annals of Operations Research*, vol. 214, no. 1, p. 167, Jul. 2011, doi: 10.1007/s10479-011-0926-z.

[16] T. Le, "The efficiency effects of bank mergers: An analysis of case studies in Vietnam," *Risk Governance and Control Financial Markets & Institutions*, vol. 7, no. 1, p. 61, Jan. 2017, doi: 10.22495/rgcv7i1art8.

[17] L. T. Vu, N. T. T. Nguyen, and L. H. Dinh, "Measuring banking efficiency in Vietnam: parametric and non-parametric methods," *Banks and Bank Systems*, vol. 14, no. 1, p. 55, Feb. 2019, doi: 10.21511/bbs.14(1).2019.06.

[18] P. H. Nguyen and D. T. B. Pham, "The cost efficiency of Vietnamese banks – the difference between DEA and SFA," *Journal of Economics and Development,* vol. 22, no. 2, p. 209, May 2020, doi: 10.1108/jed-12-2019-0075.

[19] C. Wang, N.-A.-T. Nguyen, T. Dang, and T.-T.-Q. Trinh, "A Decision Support Model for Measuring Technological Progress and Productivity Growth: The Case of Commercial Banks in Vietnam," *Axioms,* vol. 10, no. 3, p. 131, Jun. 2021, doi: 10.3390/axioms10030131.

[20] B. Kelly, S. Malamud, and K. Zhou, "The Virtue of Complexity in Return Prediction," *The Journal of Finance*, vol. 79, no. 1, p. 459, Dec. 2023, doi: 10.1111/jofi.13298.

[21] S. Giglio, B. Kelly, and D. Xiu, "Factor Models, Machine Learning, and Asset Pricing," *Annual Review of Financial Economics,* vol. 14, no. 1, p. 337, Aug. 2022, doi: 10.1146/annurev-financial-101521-104735.

[22] B. Kelly and D. Xiu, "Financial Machine Learning," *Foundations and Trends® in Finance*, vol. 13, p. 205, Jan. 2023, doi: 10.1561/0500000064.

[23] N. Gafsi, "Machine Learning Approaches to Credit Risk: Comparative Evidence from Participation and Conventional Banks in the UK," *Journal of risk and financial management,* vol. 18, no. 7, p. 345, Jun. 2025, doi: 10.3390/jrfm18070345.

[24] G. Manthoulis, M. Doumpos, C. Zopounidis, and E. Galariotis, "An ordinal classification framework for bank failure prediction: Methodology and empirical evidence for US banks," *European Journal of Operational Research*, vol. 282, no. 2, p. 786, Sep. 2019, doi: 10.1016/j.ejor.2019.09.040.

[25] P. Appiahene, Y. M. Missah, and U. Najim, "Predicting Bank Operational Efficiency Using Machine Learning Algorithm: Comparative Study of Decision Tree, Random Forest, and Neural Networks," *Advances in Fuzzy Systems,* vol. 2020, p. 1, Jul. 2020, doi: 10.1155/2020/8581202.

[26] Q. H. Nguyen, H. V. Trinh, V. P. Truong, and T. T. M. Ly, "Early Warning System for Debt Group Migration: The Case of One Commercial Bank in Vietnam," *Foundations of Management,* vol. 16, no. 1, p. 195, Jan. 2024, doi: 10.2478/fman-2024-0012.

[27] S. Boubaker, T. Le, T. Ngo, and R. Manita, "Predicting the performance of MSMEs: a hybrid DEA-machine learning approach," *Annals of Operations Research,* vol. 350, no. 2, p. 555, Feb. 2023, doi: 10.1007/s10479-023-05230-8.

[28] T. van Trung and N. A. N. Vuong, "Development Of A Credit Scoring Model Using Machine Learning For Commercial Banks In Vietnam," *Advances and Applications in Statistics*, vol. 92, no. 1, p. 107, Nov. 2024, doi: 10.17654/0972361725006.

[29] L. Yang and A. Shami, "On hyperparameter optimization of machine learning algorithms: Theory and practice," *Neurocomputing,* vol. 415, p. 295, Jul. 2020, doi: 10.1016/j.neucom.2020.07.061.

[30] B. Bischl et al., "Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges," *Wiley Interdisciplinary Reviews Data Mining and Knowledge Discovery*, vol. 13, no. 2, Jan. 2023, doi: 10.1002/widm.1484.

[31] L. Franceschi et al., "Hyperparameter Optimization in Machine Learning," *arXiv (Cornell University)*, Oct. 2024, doi: 10.48550/arxiv.2410.22854.

[32] P.-D. Arsenault, S. Wang, and J.-M. Patenaude, "A Survey of Explainable Artificial Intelligence (XAI) in Financial Time Series Forecasting," *ACM Computing Surveys*, vol. 57, no. 10. Association for Computing Machinery, p. 1, Apr. 18, 2025. doi: 10.1145/3729531.

[33] P.-D. Arsenault, S. Wang, and J.-M. Patenande, "A Survey of Explainable Artificial Intelligence (XAI) in Financial Time Series Forecasting," *arXiv (Cornell University),* Jul. 2024, doi: 10.48550/arxiv.2407.15909.

[34] B. Bilodeau, N. Jaques, P. W. Koh, and B. Kim, "Impossibility theorems for feature attribution," *Proceedings of the National Academy of Sciences,* vol. 121, no. 2, Jan. 2024, doi: 10.1073/pnas.2304406120.

ISSN 2723-6471

[35] C. Yang, M. Z. Abedin, H. Zhang, F. Weng, and P. Hájek, "An interpretable system for predicting the impact of COVID-19 government interventions on stock market sectors," *Annals of Operations Research,* vol. 347, no. 2, p. 1031, Apr. 2023, doi: 10.1007/s10479-023-05311-8.

[36] Y. Xia, Z. Liao, J. Xu, and Y. Li, "From Credit Scoring To Regulatory Scoring: Comparing Credit Scoring Models From A Regulatory Perspective," *Technological and Economic Development of Economy,* vol. 28, no. 6, p. 1954, Dec. 2022, doi: 10.3846/tede.2022.17045.

[37] P. Bracke, A. Datta, C. Jung, and S. Sen, "Machine Learning Explainability in Finance: An Application to Default Risk Analysis," *SSRN Electronic Journal,* Jan. 2019, doi: 10.2139/ssrn.3435104.

[38] F. J. Bargagli-Stoffi, F. Incerti, M. Riccaboni, and A. Rungi, "Machine learning for zombie hunting: predicting distress from firms' accounts and missing values," *Industrial and Corporate Change,* vol. 33, no. 5, p. 1063, Oct. 2023, doi: 10.1093/icc/dtad049.

[39] M. Alexandre, T. C. Silva, C. Connaughton, and F. A. Rodrigues, "The drivers of systemic risk in financial networks: a data-driven machine learning analysis," *Chaos Solitons & Fractals*, vol. 153, p. 111588, Nov. 2021, doi: 10.1016/j.chaos.2021.111588.

[40] S. Consoli, D. R. Recupero, and M. Saisana, Data Science for Economics and Finance. 2021. doi: 10.1007/978-3-030-66891-4.

[41] H. Hewamalage, K. Ackermann, and C. Bergmeir, "Forecast evaluation for data scientists: common pitfalls and best practices," *Data Mining and Knowledge Discovery*, vol. 37, no. 2, p. 788, Dec. 2022, doi: 10.1007/s10618-022-00894-5.

[42] T. Dierckx, J. Davis, and W. Schoutens, "Using Machine Learning and Alternative Data to Predict Movements in Market Risk," *arXiv (Cornell University),* Sep. 2020, doi: 10.48550/arxiv.2009.07947.

[43] V. Cerqueira, L. Torgo, and I. Mozetič, "Evaluating time series forecasting models: an empirical study on performance estimation methods," *Machine Learning,* vol. 109, no. 11, p. 1997, Oct. 2020, doi: 10.1007/s10994-020-05910-7.

[44] C. Bergmeir, R. J. Hyndman, and B. Koo, "A note on the validity of cross-validation for evaluating autoregressive time series prediction," *Computational Statistics & Data Analysis,* vol. 120, p. 70, Nov. 2017, doi: 10.1016/j.csda.2017.11.003.

[45] C. Rao, T. Xue, M. Kan, P. Zhou, and Y. Lan, "A new hybrid neural network framework inspired by biological systems for advanced financial forecasting," *Scientific Reports,* vol. 15, no. 1, p. 36985, Oct. 2025, doi: 10.1038/s41598-025-21842-5.

[46] T. Berger, "On the information content of explainable artificial intelligence for quantitative approaches in finance," *OR Spectrum*, vol. 47, no. 1, p. 177, Jun. 2024, doi: 10.1007/s00291-024-00769-9.

[47] S. Kumar, "Explainable AI in Financial Forecasting Using Time Series Analysis," *International Journal for Research in Applied Science and Engineering Technology*, vol. 13, no. 4, p. 7155, Apr. 2025, doi: 10.22214/ijraset.2025.70080.