STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–29.
Published online in International Academic Press (www.IAPress.org)

Evaluating the Credit Risk Disclosure: A Comparative Analysis of Statistical and Machine Learning Techniques for Saudi Arabian Commercial Banks

Dalia Kamal Alnagar[1], Samirah Alzubaidi [2], Uzma Yasmeen[3,4], Amr R. Kamel[5*], Hazim G. Kalt[6], M.S. Mohamed[7], Ahood Alazwari[8], Eslam Hussam[9], Ezzat Hussein Khmar[10], Zaid Mohammed Alqahtani[11]

[1] *Department of Statistics, Faculty of Science, University of Tabuk, Tabuk, Saudi Arabia*
[2] *Department of Mathematics, Al-Qunfudah University College, Umm Al-Qura University, Mecca, Saudi Arabia*
[3] *Institute of Molecular Biology and Biotechnology (IMBB), The University of Lahore,Lahore, Pakistan*
[4] *Department of Mathematics & Statistics, Brock University, Saint Catherines Ontario, Canada*
[5] *Department of Applied Statistics and Econometrics, Faculty of Graduate Studies for Statistical Research, Cairo University, Giza, Egypt*
[6] *Department of Mathematics, College of Education for Pure Science, University of Kerbala, Iraq*
[7] *Department of Mathematics, College of Science and Humanities, Prince Sattam bin Abdulaziz University, Hawtat Bani Tamim 16511, Saudi Arabia*
[8] *Department of Mathematics, Faculty of Science, Al-Baha University, Al-Baha, Saudi Arabia*
[9] *Department of Mathematics, Faculty of Science, Capital University, Cairo, Egypt*
[10] *Department of Statistics, Faculty of Commerce, Al-Azhar University (Boys' Branch), Egypt*
[11] *Department of Accounting, College of Business Administration, University of Hafr Al-Batin, Saudi Arabia*

Abstract This study examines whether reported credit-risk and bank-specific indicators can predict the financial performance of Saudi Arabian commercial banks using machine-learning (ML) techniques. The empirical analysis is based exclusively on actual secondary financial data collected from the published annual reports of ten Saudi Arabian commercial banks over the period 2015–2025, yielding 110 bank-year observations. Return on Assets (ROA), Return on Equity (ROE), and Earnings per Share (EPS) are used to measure financial performance, and several bank-specific and credit-risk indicators are used as explanatory factors. Financial performance is classified using four ML techniques: SVM, XGBoost, ANN, and KNN. The prediction performance of these techniques is assessed using Accuracy, Precision, Recall, F1-Score, and AUC-ROC. With accuracy rates of 78.79% and 84.85% and AUC values of 0.838 and 0.864, respectively, the results show that SVM offers the most reliable overall predictive performance, especially for ROA and ROE. KNN has the best predictive performance for EPS, with an accuracy of 78.79%, an F1-Score of 81.08%, and an AUC of 0.816. The feature-importance analysis also shows that two of the most important variables in predicting financial performance are capital adequacy and financial efficiency. With SVM appearing as the most dependable approach overall, the results show the value of ML approaches in spotting intricate patterns and enhancing the forecasting of financial performance in Saudi Arabian commercial banks.

Keywords Credit Risk Indicators, Financial Performance, Saudi Arabian Commercial Banks, Machine Learning Techniques, AUC-ROC, Predictive Analytics.

AMS 2010 subject classifications 62P05, 62H30, 68T05, 91G40

DOI: 10.19139/soic-2310-5070-4535

*Correspondence to: Amr R. Kamel (Email: amr_ragab@pg.cu.edu.eg). Department of Applied Statistics and Econometrics, Faculty of Graduate Studies for Statistical Research, Cairo University, Giza, Egypt.

ISSN 2310-5070 (online) ISSN 2311-004X (print)
Copyright © 202x International Academic Press

1. Introduction

In Saudi Arabian commercial banks, credit risk disclosure is an essential part of financial reporting since it gives stakeholders clear information about the banks' exposure to borrower default risk and their capacity to handle possible credit losses. The main ways that Saudi Arabian commercial banks are exposed to credit risk are through interbank placements, debt securities, client loans and advances, and other credit-related financial instruments. Therefore, thorough information about the type and extent of credit exposures, the quality and concentration of the loan portfolio, non-performing loans, credit-impaired assets, provisions and anticipated credit losses, collateral and guarantees, and the policies and procedures used to identify, measure, monitor, and control credit risk should all be included in effective credit risk disclosure [1]. Given changes in macroeconomic conditions, interest rates, exchange rates, and borrowers' financial capacity, all of which can have an impact on the quality of banks' loan portfolios, such disclosure is especially crucial in the Saudi Arabian banking industry. Investors, depositors, regulators, and other stakeholders can evaluate the effectiveness of banks' risk-management procedures and financial resilience thanks to transparent credit risk reporting, which also improves the dependability and comparability of financial statements and fortifies market discipline.

The potential losses that could occur if a counterparty (obligor or bank) to the Saudi Central Bank does not fulfil a financial obligation for full or partial value on the due date are known as credit risk. Within the scope of the approved Saudi Central Bank credit policy, which establishes the rules, guidelines, and principles used to manage these risks and limit them to the lowest possible level, the Credit Risk Department works to mitigate such risk [2]. It also outlines each party's obligations and responsibilities under this framework. In order to determine credit limits for the central bank's counterparties, the Credit Risk Department conducts thorough quantitative and qualitative analysis in addition to sovereign risk analysis. These restrictions apply to trade finance transactions and Nostro accounts, which are handled by the Banking Operations Sector of the Saudi Central Bank, as well as to money market placements and foreign exchange transactions, which are overseen by its Reserve Management Department. Additionally, the Credit Risk Department monitors and investigates the solvency of foreign securities traders.

The Credit Risk Department uses the central bank's Obligor Risk Rating (ORR), bank quantitative and qualitative analysis, and country risk (political, economic, fiscal, and banking sectors) to determine credit limits. The three major international rating agencies (Fitch, Standard & Poor's, and Moody's) assign an ORR to the counterparties of the Saudi Central Bank based on their lowest rating grade [3]. Additionally, the Credit Risk Department keeps a close eye on events that could affect the central bank's counterparties and the nations in which they are based, in order to spot any developments that might raise the credit risk to which the Saudi Central Bank is exposed and take the necessary action.

Several theoretical foundations can be used to explain the connection between bank financial success and credit-risk indicators. According to agency theory, open indicators strengthen accountability and supervision by reducing the information asymmetry between bank management and stakeholders. According to signaling theory, thorough indicators conveys to depositors and investors the bank's capacity for risk management and financial stability. Additionally, according to legitimacy theory, banks should reveal pertinent risk information to uphold stakeholder confidence and show that they are adhering to institutional and regulatory requirements. These theoretical stances offer a basis for comprehending the significance of credit-risk indicators in the Saudi banking sector.

The primary aspects of credit-risk disclosure in Saudi Arabian commercial banks are depicted in Figure 1, which also emphasises the contribution of disclosure to financial stability, accountability, and transparency. Credit risk disclosure, which includes details about the type and scope of credit exposures, loan portfolio quality and concentration, credit-impaired assets and non-performing loans, provisions and anticipated credit losses, collateral and guarantees, credit risk management procedures, risk measurement, and internal rating systems, is at the core of the framework. These factors give interested parties a thorough grasp of the degree and handling of credit risk in Saudi Arabian commercial banks. The figure also shows how several external factors, such as macroeconomic conditions, interest rate and exchange rate variations, borrowers' financial ability, and the regulatory and supervisory environment, affect credit risk in the Saudi Arabian banking sector. Effective disclosure should therefore go beyond simply providing credit exposure data and include an explanation of the policies,

practices, and measurement methods that banks employ to detect, track, and reduce credit risk. Users can assess the possible susceptibility of banks to credit deterioration by paying close attention to non-performing loans, predicted credit losses, collateral, and internal credit ratings.

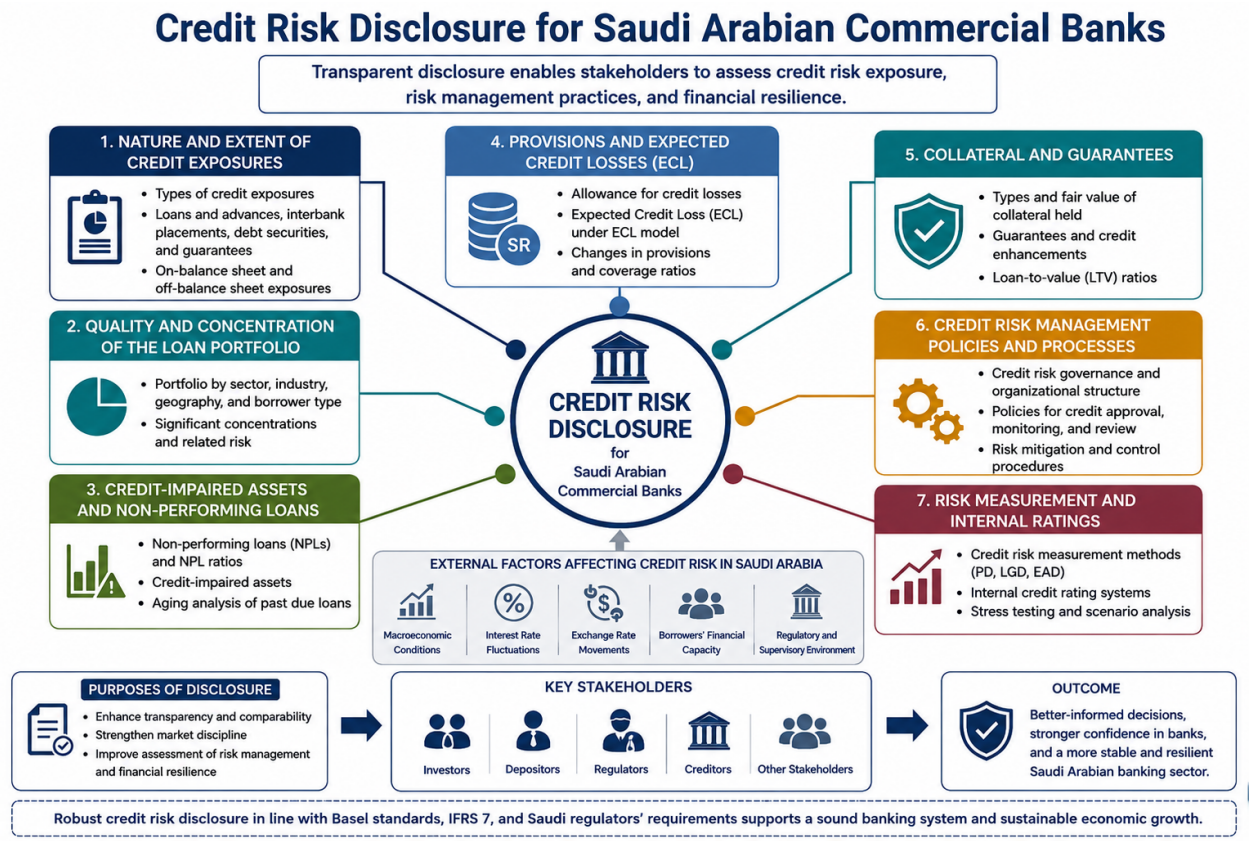


Figure 1. Main dimensions of credit risk disclosure in Saudi Arabian commercial banks.

Finally, the figure highlights how crucial credit risk disclosure is for regulators, creditors, depositors, investors, and other stakeholders. Good disclosure helps stakeholders evaluate banks' financial resilience, promotes informed decision-making, strengthens market discipline, and boosts trust in the banking industry. Therefore, thorough credit risk disclosure can support improved risk governance, increased transparency, and a more robust and stable commercial banking industry in Saudi Arabia.

The study uses credit-risk indicators reported in banks' published financial statements (e.g., NPLs, loan loss provisions, capital adequacy) as predictors of financial performance. Disclosure is discussed as the institutional context in which these indicators are reported, not as a measured variable. The contribution of the paper is to compare the predictive ability of four ML techniques for this purpose.

This paper is structured as follows. Section 2 presents a comprehensive literature review, outlining credit risk disclosure with different analysis models. Section 3 provides a detailed content analysis of machine learning techniques and algorithms. Section 4 discusses methodological innovations, data collection, and sample selection. Section 5 offers a broader discussion of results. Finally, Section 6 concludes the paper by synthesising the main insights and drawing some conclusions.

2. Literature Review

A thorough analysis of credit risk determinants is necessary to determine the most important factors endangering the survival and expansion of banks in the economy, as macroeconomic and bank-specific variables have a substantial impact on credit risk levels within banks [4]. In this regard, key macroeconomic factors influencing credit risk were identified by Radivojević et al. [5], Niepmann and Schmidt-Eisenlohr [6], and Abinzano et al. [7]. These factors include economic growth, exchange rates, interest rates, inflation, and public and external debt. Therefore, it is important for banks to regularly monitor these variables to improve their ability to predict credit risk exposure. Furthermore, bank-specific metrics including capital sufficiency, income diversification, operating efficiency, liquidity, asset size, and profitability are crucial for predicting future credit risk levels, as demonstrated by Barra and Ruggiero [8]. They discovered a negative relationship between credit risk and economic growth as measured by the real Gross Domestic Product (GDP) growth rate. This implies that increased economic growth strengthens the financial stability of businesses in a variety of industries, increasing their ability to repay debt and lowering their exposure to credit risk.

Additionally, academics have been debating the detrimental effects of public debt on credit risk, pointing out that greater government budget deficits require the issuance of more Treasury securities to make up for revenue shortfalls [7]. As a result, banks might shift their reserves into government securities, thereby reducing their exposure to credit risk and loan investments. On the other hand, Ofria and Mucciardi [19] and Naili and Lahrichi [4] discovered that public debt may have a favourable impact on bank credit risk. They contended that a rise in the issuance of Treasury securities boosts market interest rates, reduces total bank reserves, and contracts the money supply, all of which raise financing costs for individuals and companies and increase banks' exposure to credit risk. Table 1 presents a summary of selected studies on credit risk disclosure.

Table 1. Summary of literature studies on credit risk disclosure.

Study	**Context / Sample**	**Main Objective**	**Method**	**Key Findings**
Linsley and Crumpton [9]	UK & Canadian banks	Examined the risk disclosure practices.	Content analysis	Risk disclosure varied considerably, with limited quantitative and forward-looking information.
Elbannan and Elbannan [10]	Egyptian banks	Examined the economic consequences of bank risk disclosure.	Disclosure analysis & regression	Risk disclosure was positively associated with operating performance and market value.
Nahar et al. [11]	Bangladeshi banks	Identify determinants of risk disclosure.	Content analysis & regression	IFRS 7 and Basel II implementation increased the level of risk disclosure.
Al-Maghzom et al. [12]	Saudi commercial banks	Examined the effect of credit risk disclosure on financial performance.	Empirical analysis	Credit risk disclosure was important for reducing credit risk and improving financial performance.
Abd Elghaffar et al. [13]	Egyptian banks	Identify determinants of bank risk disclosure.	Content analysis & regression	Disclosure was affected by corporate governance, board characteristics, audit committee size, ownership, and bank characteristics.
Scannella and Polizzi [14]	Italian banks	Measure the quality of credit risk disclosure.	Content analysis	Significant differences existed in credit risk disclosure among banks despite similar regulatory requirements.
Qadiri and Alsughayer [15]	Saudi banks	Examined credit risk disclosure and its determinants.	Disclosure index & regression	Bank size and leverage were important determinants of credit risk disclosure.
Hussein [16]	Egyptian commercial banks	Examined the effect of credit risk disclosure on financial performance.	Regression analysis	Credit risk disclosure had a significant effect on banks' financial performance.
Altunbaş et al. [17]	European banks	Examine the relationship between risk disclosure and bank risk.	Text analysis & regression	Greater disclosure was generally associated with lower bank default risk.
Lobo et al. [18]	53 countries	Examined the effect of IFRS 7 adoption on bank risk-taking.	Difference-in-differences	The adoption of IFRS 7 reduced banks' risk-taking and encouraged more conservative lending practices.

The studies in Table 1 measure disclosure using content analysis or disclosure indices. The present study differs: it relies on quantitative credit-risk indicators rather than a disclosure score. Hence, it contributes to the literature on credit-risk determinants and ML-based performance prediction, and it complements, rather than replaces, the disclosure-index literature.

The application of ML algorithms to textual risk information is a significant development in this literature. One particularly pertinent study investigated whether actual credit risk might be explained by the soft data found in the annual reports of European banks. In order to convert narrative data from yearly reports into a text-based credit-risk metric, the researchers used supervised ML algorithms. They discovered that the ML-derived metric accounted for a significant amount of variation in NPLs, Ohlson's O-score, Altman's Z-score, and credit-rating downgrades using 1,885 bank-year records from 19 European nations between 2005 and 2017. The study shows that narrative disclosures can yield economically significant information that traditional financial ratios can miss [20].

Recent studies have highlighted the growing application of ML techniques in financial risk assessment. Seliem et al. [21] compared several supervised learning algorithms for credit risk prediction and examined the role of feature selection in improving classification performance. Elsegai et al. [22] compared several machine learning and deep learning techniques for predicting stock market trends, including SVM, KNN, ANN, Random Forest, and AdaBoost. Their findings illustrate the importance of comparing alternative predictive algorithms in financial applications. Although their study addresses stock market prediction rather than bank profitability, it provides a relevant methodological foundation for the comparative machine learning framework adopted in this study. Similarly, Faris and Elhachloufi [23] investigated various artificial intelligence and ML models for credit risk prediction using real-world financial data. Hjouji et al. [24] proposed an ML approach specifically designed to predict credit risk in bank loans. Building on this literature, the present study extends the empirical application of ML techniques to the prediction of financial performance among Saudi Arabian commercial banks, focusing on Return on Assets (ROA), Return on Equity (ROE), and Earnings per Share (EPS).

The reviewed studies show that credit risk disclosure has drawn a lot of attention from around the world, but there is still a dearth of evidence from Saudi Arabian commercial banks, especially when it comes to the creation of a comprehensive credit risk disclosure index and the identification of its determinants. Previous research conducted in Saudi Arabia has mostly focused on general risk disclosure or the connection between financial success and credit risk disclosure. Thus, by offering a targeted evaluation of credit risk disclosure among Saudi Arabian commercial banks and identifying the variables linked to its level, the current study aims to expand the body of literature.

Although previous studies have examined bank risk disclosure, limited empirical attention has been devoted to comparing machine-learning classification techniques for evaluating the relationship between credit-risk indicators and financial performance in Saudi Arabian commercial banks. Accordingly, this study contributes to the literature by providing a comparative predictive assessment of four ML classifiers: Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), Artificial Neural Networks (ANN), and K-Nearest Neighbours (KNN), using bank-specific and credit-risk indicators to classify financial performance. Its contribution lies in the comparative empirical application and evaluation of these techniques within the Saudi banking context rather than proposing a new classification algorithm or constructing a credit-risk disclosure index. To address the identified research gap, this study seeks to answer the following research questions:

RQ1: To what extent can credit-risk indicators predict the financial performance of Saudi Arabian commercial banks, measured by ROA, ROE, and EPS?
RQ2: Which credit-risk indicators provide the greatest predictive importance in classifying banks into high- and low-financial-performance groups?
RQ3: Which ML techniques: SVM, XGBoost, ANN, or KNN provide the strongest predictive performance?
RQ4: Do the predictive performances of these techniques differ across the three financial performance measures?

Based on the reviewed literature, credit-risk indicators are expected to contain useful predictive information about banks' financial performance. However, their relative importance may differ across ROA, ROE, and EPS. Furthermore, machine-learning algorithms may exhibit different predictive capabilities because of differences in their ability to capture nonlinear patterns and complex relationships. Therefore, this study empirically compares four machine-learning techniques without assuming in advance that one technique will outperform all others.

3. Machine Learning Techniques

ML techniques have become increasingly important in empirical research because they provide powerful tools for identifying complex patterns, relationships, and nonlinear associations within large and multidimensional datasets. ML techniques can discover patterns directly from the observed data and enhance predictive performance through systematic training, in contrast to classic statistical approaches that frequently rely on predetermined functional forms and assumptions about the relationships among variables [25], [26], [27] and [28]. Elbatal et al. [29] investigated the application of traditional statistical forecasting methods and machine learning techniques to financial accounting data. Their comparison demonstrated the potential of multilayer perceptron models to improve sales forecasting accuracy. This provides additional motivation for evaluating flexible machine learning approaches alongside conventional statistical models in financial prediction problems. Credit risk assessment, bankruptcy prediction, fraud detection, financial distress prediction, and disclosure analysis are just a few of the areas in accounting, finance, and banking research where machine learning approaches are being used more and more. They are especially helpful for forecasting financial and banking outcomes because of their ability to handle high-dimensional datasets, non-linear correlations, and interactions among explanatory variables [30]. SVM, XGBoost, ANN, and KNN are among the most widely used methods. Figure 2 shows the types of ML techniques. Assume that the Saudi banks dataset is

$$D = \{(\mathbf{x}_i, y_i) : i = 1, \ldots, n\}, \tag{1}$$

where $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^\top$ is the explanatory variable vector, and y represents the associated financial-performance class. Estimating a function $f : \mathbb{R}^p \to y$ is the main goal of supervised ML techniques, which reduces the anticipated loss in prediction:

$$f^* = \arg\min_{f \in \mathcal{F}} \; E\big[L\big(Y, f(X)\big)\big]\,. \tag{2}$$

The empirical risk approximates the predicted risk when the population distribution is unknown:

$$\widehat{R}(f) = \frac{1}{n} \sum_{i=1}^{n} L\big(y_i, f(\mathbf{x}_i)\big). \tag{3}$$

A regularisation term might be added to lessen overfitting:

$$\widehat{f} = \arg\min_{f \in \mathcal{F}} \left[\frac{1}{n} \sum_{i=1}^{n} L\big(y_i, f(\mathbf{x}_i)\big) + \lambda\, \Omega(f) \right], \tag{4}$$

where λ regulates the trade-off between prediction accuracy and complexity, $\Omega(f)$ measures model complexity, and $L(\cdot)$ is the loss function.

Supervised, unsupervised, and reinforcement learning are the three primary categories of machine learning. In supervised learning, models are trained using labelled data for both classification and regression. Finding patterns and clustering in unlabelled data are the main goals of unsupervised learning. By interacting with the environment, reinforcement learning enhances model performance.

3.1. Support Vector Machine

A supervised machine learning method called Support Vector Machine (SVM) can be applied to both regression and classification problems. Finding an ideal decision boundary, or hyperplane, that divides data from various classes while maximising the distance between the border and the nearest observations, known as support vectors, is the core goal of SVM. Using kernel functions to handle high-dimensional data and nonlinear interactions is one of SVM's main advantages [31]. Statistical learning theory, structural risk minimisation, and margin maximisation form the mathematical basis of SVM. Let $y \in \{-1, +1\}$ for binary categorisation. The definition of a linear separating hyperplane is

$$f(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + b. \tag{5}$$

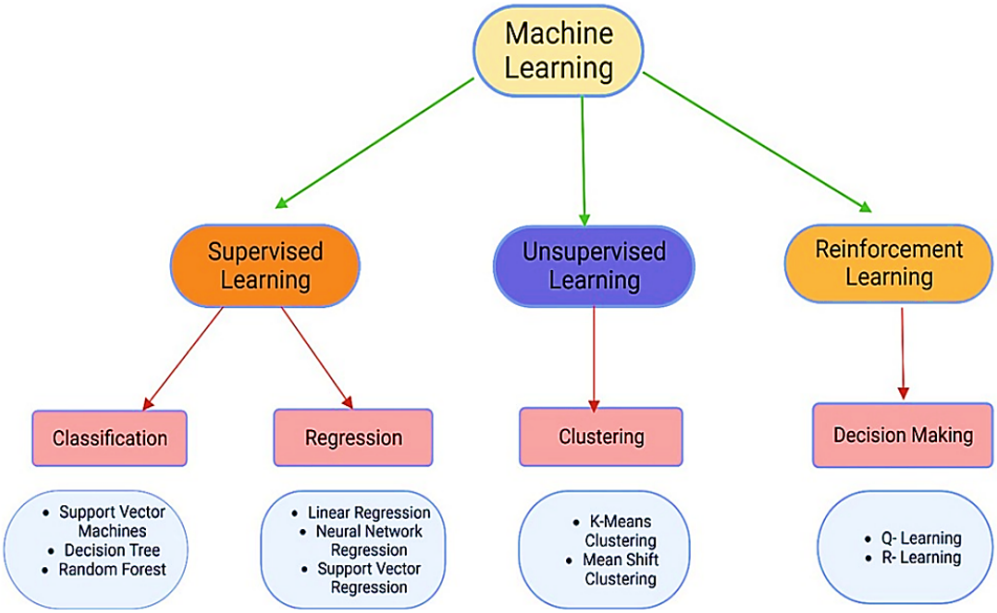


Figure 2. Types of machine learning techniques.

The classification rule is

$$\widehat{y} = \operatorname{sign}\left(\mathbf{w}^\top \mathbf{x} + b\right). \tag{6}$$

The distance of an observation is

$$d_i = \frac{\left|\mathbf{w}^\top \mathbf{x}_i + b\right|}{\|\mathbf{w}\|}. \tag{7}$$

The distance between the two classes is maximised via SVM. Conversely, it reduces

$$\min_{\mathbf{w},b} \ \tfrac{1}{2}\|\mathbf{w}\|^2, \qquad \text{subject to} \qquad y_i\left(\mathbf{w}^\top \mathbf{x}_i + b\right) \geq 1.$$

Slack variables $\xi_i \geq 0$ are introduced when perfect separation is not possible:

$$\min_{\mathbf{w},b,\boldsymbol{\xi}} \ \tfrac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n} \xi_i, \tag{8}$$

subject to

$$y_i\left(\mathbf{w}^\top \mathbf{x}_i + b\right) \geq 1 - \xi_i, \qquad \xi_i \geq 0. \tag{9}$$

The matching hinge loss is

$$L\big(y_i, f(\mathbf{x}_i)\big) = \max\left\{0,\ 1 - y_i f(\mathbf{x}_i)\right\}. \tag{10}$$

Thus, the SVM objective can be written as

$$\min_{\mathbf{w},b} \left[\tfrac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n} \max\left(0,\ 1 - y_i f(\mathbf{x}_i)\right)\right].$$

The penalty for categorisation errors is determined by the parameter C. SVM can be extended to nonlinear relationships using a kernel function:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \varphi(\mathbf{x}_i)^\top \varphi(\mathbf{x}_j). \tag{11}$$

For non-linear classification, regarding the kernel of the radial basis function,

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2\right). \tag{12}$$

Because SVM can detect nonlinear boundaries between various levels of bank financial performance while regulating model complexity through margin maximisation and regularisation, it is theoretically appealing for the current investigation.

When the link between explanatory variables and the dependent variable is complicated and challenging to describe using traditional linear models, SVM can be very helpful in financial and banking research. By regulating model complexity with parameters such as the regularisation parameter C and kernel-specific parameters, the method also offers strong generalisation performance [32]. As a result, SVM can be used to anticipate outcomes related to financial and credit risk as well as to categorise institutions based on their risk or disclosure levels.

3.2. Extreme Gradient Boosting

Based on the gradient boosting framework, Extreme Gradient Boosting (XGBoost) is a sophisticated ensemble machine learning technique. It builds a series of decision trees, each of which tries to fix the prediction errors produced by the trees before it. In contrast to traditional single-tree models, XGBoost creates a powerful predictive model by combining the predictions of numerous weak learners. The method's computational efficiency makes it appropriate for reasonably big and complicated datasets, and it features regularisation procedures that assist in reducing overfitting and increasing generalisation [33]. Additionally, XGBoost can capture interactions and nonlinear correlations between explanatory factors without the researcher's explicit specification.

XGBoost builds an ensemble of decision trees sequentially, with each tree minimising the prediction error of the previous trees:

$$\widehat{y}_i = \sum_{k=1}^{K} f_k(\mathbf{x}_i), \qquad f_k \in \mathcal{F}. \tag{13}$$

The objective function is

$$\mathcal{L}(\phi) = \sum_{i=1}^{n} l\left(y_i, \widehat{y}_i\right) + \sum_{k=1}^{K} \Omega(f_k), \tag{14}$$

where the regularisation term is

$$\Omega(f_k) = \gamma T + \tfrac{1}{2}\lambda \|w\|^2. \tag{15}$$

Thus, the prediction is predicated on the total contribution of all trees:

$$\widehat{y}_i^{(t)} = \widehat{y}_i^{(t-1)} + f_t(\mathbf{x}_i). \tag{16}$$

A second-order Taylor expansion of the loss function is employed by XGBoost:

$$l\left(y_i, \widehat{y}_i^{(t-1)} + f_t(\mathbf{x}_i)\right) \approx l\left(y_i, \widehat{y}_i^{(t-1)}\right) + g_i f_t(\mathbf{x}_i) + \tfrac{1}{2} h_i f_t^2(\mathbf{x}_i), \tag{17}$$

where

$$g_i = \frac{\partial\, l\left(y_i, \widehat{y}_i^{(t-1)}\right)}{\partial\, \widehat{y}_i^{(t-1)}} \qquad \text{and} \qquad h_i = \frac{\partial^2 l\left(y_i, \widehat{y}_i^{(t-1)}\right)}{\partial \left(\widehat{y}_i^{(t-1)}\right)^2}.$$

The new tree can be efficiently optimised thanks to this second-order approximation. As a result, XGBoost provides a solid mathematical foundation for both controlling overfitting through regularisation and capturing nonlinear effects, interactions, and intricate linkages between credit-risk indicators and financial performance.

Credit risk, financial hardship, default probability, and other categorical or continuous financial outcomes can all be predicted using XGBoost in banking and financial research. Additionally, XGBoost is especially helpful for empirical research where both prediction and interpretation are crucial, since explainability techniques and variable-importance measures can be used to identify the factors that contribute most significantly to the model's predictions [34].

3.3. Artificial Neural Network

Computational models called Artificial Neural Networks (ANN) draw inspiration from the architecture and information-processing techniques of biological neural networks. An input layer, one or more hidden layers, and an output layer typically make up an ANN. After applying weights and an activation function to the data it receives from the preceding layer, each neuron sends the resulting data to the following layer [35]. The network uses backpropagation and optimisation methods to modify its weights during training in order to minimise the discrepancy between expected and observed results. ANN consists of interconnected neurons organised into input, hidden, and output layers. The input of a neuron is

$$z_j = \sum_{i=1}^{p} w_{ij} x_i + b_j. \tag{18}$$

The activation function is

$$h_j = \varphi(z_j), \tag{19}$$

and for a sigmoid activation function,

$$\varphi(z) = \frac{1}{1+e^{-z}}. \tag{20}$$

The output probability for binary classification is

$$\widehat{p}_i = \frac{1}{1+\exp(-z_i)}.$$

The loss function for binary cross-entropy is

$$L = -\frac{1}{n}\sum_{i=1}^{n} \left[y_i \log(\widehat{p}_i) + (1-y_i)\log(1-\widehat{p}_i) \right]. \tag{21}$$

The estimation of neural network parameters is done by

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} L(\boldsymbol{\theta}), \qquad \text{and} \qquad \boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - \eta \, \nabla_{\boldsymbol{\theta}} L,$$

where η is the learning rate. The derivative of the sigmoid activation is

$$\varphi'(z) = \varphi(z)\left[1-\varphi(z)\right], \tag{22}$$

which makes the backpropagation process easier. ANN is capable of approximating very nonlinear relationships through repeated forward and backward propagation:

$$f(\mathbf{x}) \approx f^*(\mathbf{x}). \tag{23}$$

Because ANN may approximate complex nonlinear mappings between credit-risk information and financial-performance outcomes without imposing a preconceived functional shape, it is theoretically significant in the current work. The commonly used mean squared error loss function is

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n} (y_i - \widehat{y}_i)^2. \tag{24}$$

ANN's capacity to represent intricate and highly nonlinear interactions between variables without requiring a particular mathematical functional form is one of its main advantages. Because of this, ANN is especially suitable for financial datasets where there may be nonlinear correlations between bank attributes, risk indicators, governance variables, and disclosure results [36]. However, ANN typically needs adequate training data and proper parameter adjustment, and its intricate internal structure may make its predictions harder to understand than those of more straightforward statistical models.

3.4. K-Nearest Neighbours

A non-parametric supervised learning method called K-Nearest Neighbours (KNN) uses the features of the nearest observations in the training dataset to classify or predict an observation. The technique finds the K observations that are most comparable to the new observation by calculating the distance between observations, usually using Euclidean distance. The majority class of the chosen neighbours usually determines the anticipated class for classification problems, while the average values of the nearby observations can be used to determine the forecast for regression problems [37]. KNN predicts the outcome of an observation based on the closest observations in the training dataset. The Euclidean distance between two observations is

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{\sum_{l=1}^{p} (x_{il} - x_{jl})^2}. \tag{25}$$

For classification, the predicted class is

$$\widehat{y}(\mathbf{x}) = \text{mode}\left\{y_i : \mathbf{x}_i \in N_K(\mathbf{x})\right\}. \tag{26}$$

For regression, the predicted value is

$$\widehat{y}(\mathbf{x}) = \frac{1}{K} \sum_{\mathbf{x}_i \in N_K(\mathbf{x})} y_i, \tag{27}$$

where $N_K(\mathbf{x})$ represents the K nearest observations to $\mathbf{x}$.

Strong assumptions on the underlying distribution of the data are not necessary for KNN, and it is comparatively easy to implement. As a result, it can be helpful in banking and financial applications where there may not be a preset functional form for the link between explanatory variables and the result. However, KNN's processing requirements may rise significantly as the amount of data grows, and it can be sensitive to the scale of variables and the selection of K. Because of this, when using KNN on banking and financial information, proper data standardisation and parameter selection are crucial [38]. Table 2 provides a summary of the mathematical underpinnings of the ML techniques.

Table 2. Theoretical evaluation of ML techniques using their mathematical foundations.

Technique	**Mathematical Foundation**	**Optimisation / Decision Principle**	**Main Strength**
SVM	Statistical learning theory	Margin maximisation + regularisation	Robust nonlinear classification
XGBoost	Gradient boosting theory	First- and second-order loss optimisation	Complex nonlinearities and interactions
ANN	Nonlinear approximation theory	Gradient-based optimisation	Highly nonlinear relationships
KNN	Nonparametric learning theory	Local distance-based classification	Flexible local patterns

As a result, the four methods offer complementary mathematical representations of the same prediction problem. SVM prioritises maximum-margin classification, XGBoost minimises a regularised ensemble loss, ANN uses iterative optimisation to estimate a nonlinear function, and KNN uses distances to achieve local nonparametric classification. Because the underlying relationships may entail nonlinearities, interactions, and varied patterns that may not be sufficiently reflected by a single functional form, this theoretical variety is especially pertinent to the analysis of credit risk and bank financial performance.

3.5. Assessment Criteria

The performance ranking of ML algorithms in our study is evaluated using four measurement criteria: accuracy, precision, recall, and F1-score. It is commonly believed that the best way to assess the performance of classification

algorithms is by using evaluation measures. The statistical metrics are calculated using the following formulas:

$$\text{Precision} = \frac{\text{TRPO}}{\text{TRPO} + \text{FANE}} \times 100, \tag{28}$$

$$\text{Recall} = \frac{\text{TRPO}}{\text{TRPO} + \text{FAPO}} \times 100, \tag{29}$$

$$\text{F1-Score} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \times 100, \tag{30}$$

$$\text{Accuracy} = \frac{\text{TRPO} + \text{TRNE}}{\text{Total number of samples}} \times 100, \tag{31}$$

where a true positive (TRPO) means the model predicted a positive outcome, and indeed it is positive. A false positive (FAPO) is when the model predicts a positive outcome, but the actual outcome is negative. A true negative (TRNE) means the model predicted a negative outcome, and it is indeed negative. A false negative (FANE) is when the model predicts a negative outcome, but the actual outcome is positive [39]. The area under the receiver operating characteristic curve (AUC-ROC) was designated as the primary evaluation metric because it measures the ability of a classifier to distinguish between high- and low-financial-performance classes across different classification thresholds. Accuracy, precision, recall, and F1-score were used as complementary metrics to provide a more comprehensive assessment of classification performance. Therefore, model selection was based primarily on AUC-ROC, while the remaining metrics were considered to identify differences in predictive performance and classification errors.

4. Data Collection and Methodology

The Saudi Central Bank, which sets regulatory and supervisory standards for capital adequacy, credit-risk management, financial stability, and banking transparency, is in charge of overseeing the Saudi banking industry. Additionally, Basel regulatory standards prioritize effective risk management, capital sufficiency, and transparency, whereas IFRS 7 introduces disclosure requirements about the nature and amount of financial-instrument risks, including credit risk. Credit-risk data is especially important when assessing the stability and financial performance of Saudi commercial banks because of these legal obligations.The explanatory variables employed in this study are financial indicators reflecting banks' credit-risk exposure and financial characteristics; they are not direct measures of the extent or quality of credit-risk disclosure. These variables were extracted from the published annual financial reports of the sampled banks and calculated according to their respective definitions. Therefore, the study focuses on predicting financial performance using credit-risk-related financial indicators rather than measuring compliance with credit-risk disclosure requirements.

The study population consists of banks in the public sector and the private sector (commercial and specialised banks) affiliated with the Saudi Central Bank, which disclosed their financial positions and closing accounts for the period 2015–2025, comprising a total of 39 banks; see Table 3. As of September 2025, 39 banks were licensed by the Saudi Central Bank: 11 local banks, 4 digital banks, and 24 foreign branches. The research sample was selected using stratified sampling and consists of 10 banks, representing 26% of the total study population, with 110 observations. The data were collected from the financial reports of the banks under study.[†] These banks were selected based on their total asset size, paid-up capital, and the amount of net loans granted to customers, as well as their widespread presence in the Saudi Arabian banking market at the national level. All observations, including those for 2025, are drawn from the complete audited annual financial statements of the sampled banks for the fiscal year ending 31 December 2025. No interim, quarterly, or annualised figures were used, and no observations were imputed or dropped.

[†]Saudi Central Bank, https://www.sama.gov.sa/en-US/Pages/default.aspx.

Table 3. Structural distribution of licensed banks in the Kingdom of Saudi Arabia.

Bank Category	**Count**	**Banking Details**
Local Commercial Banks	11	National conventional and Islamic banks (e.g., Al Ahli, Al Rajhi, Riyad, Al Awwal, Al Inma, Al Bilad, etc.).
Local Digital Banks	4	Fully digital banks without traditional branches: STC Bank, D360 Bank, Vision Bank, and EZ Bank.
Foreign Bank Branches	24	Licensed branches of GCC and international banks operating within the Saudi market.
Total	**39**	**All licensed and registered banks with the Saudi Central Bank.**

Table 4. Indicators and measurement methods of the variables.

No.	**Indicator**	**Measurement Method**	**Code**
Response Variables			
1	Return on Assets	Net profit after tax / total assets	ROA
2	Return on Equity	Net profit after tax / shareholders' equity	ROE
3	Earnings per Share	Net profit / weighted average number of ordinary shares	EPS
Explanatory Variables			
4	Non-Performing Loans Ratio	Total non-performing loans / total loans and advances	NPLR
5	Loan Loss Provision Ratio	Loan loss provisions / total loans and advances	LLPR
6	Loans and Advances to External and Internal Resources Ratio	Total loans and advances / (total deposits + shareholders' equity)	LEIR
7	Shareholders' Equity to Total Loans Ratio	Shareholders' equity / total loans	SETR
8	Liquidity Ratio	Total loans / total deposits	LIRA
9	Financial Leverage Ratio	Total short-term and long-term liabilities / shareholders' equity	FILR
10	Financial Efficiency Ratio	Net profit after tax / total revenues	FIER
11	Capital Adequacy Ratio	Total shareholders' equity / total assets	CAAR
12	Bank Size	Natural logarithm of total bank assets at the end of each financial year	BASI

The purpose of this study is to investigate how credit risk disclosure affects the financial performance of Saudi Arabian commercial banks. As indicated in Table 4, the primary explanatory variable is credit risk disclosure and the dependent variable is financial performance. To give a thorough evaluation of this link, the analysis makes use of a number of credit risk and bank-specific indicators. The Non-Performing Loans Ratio (NPLR), Loan Loss Provision Ratio (LLPR), Loans and Advances to External and Internal Resources Ratio (LEIR), Shareholders' Equity-to-Total Loans Ratio (SETR), Liquidity Ratio (LIRA), Financial Leverage Ratio (FILR), Financial Efficiency Ratio (FIER), Capital Adequacy Ratio (CAAR), and Bank Size (BASI) are the explanatory variables. Three well-known metrics, Return on Assets (ROA), Return on Equity (ROE), and Earnings per Share (EPS), are used to gauge financial performance. These factors enable the study to determine whether variances in the financial performance of Saudi Arabian commercial banks are linked to differences in credit risk exposure, capital structure, liquidity, efficiency, and bank size. Several ratios are mechanically related to

the outcomes: ROE = ROA × (assets/equity), CAAR (equity/assets) is the inverse of this leverage term, FIER (net profit/revenue) is the margin component of ROA, and FILR, SETR and LEIR overlap through balance-sheet identities. High importance for these variables therefore partly reflects accounting identities, not independent information.

The study uses four ML techniques – SVM, XGBoost, ANN, and KNN – to investigate the connection between credit risk disclosure and the financial performance of Saudi Arabian commercial banks. Using 110 bank-year observations from ten Saudi Arabian commercial banks, the research includes three financial performance metrics (ROA, ROE, and EPS), as well as nine explanatory factors that characterise credit risk and bank-specific traits. Based on the explanatory variables, each algorithm is trained to recognise and categorise the banks' financial performance levels. Precision, Recall, F1-Score, Accuracy, and the confusion matrix are then used to evaluate the predictive performance of the four models. The best ML method for evaluating and forecasting the financial performance of Saudi Arabian commercial banks in connection with credit risk disclosure is then determined by comparing the outcomes of the four models. Figure 3 shows the general structure for analysis and evaluation.

To ensure a fair comparison and reproducibility of the machine learning models, a systematic preprocessing and hyperparameter-tuning procedure was adopted. All predictor variables were standardized using parameters estimated from the training data, and the same transformations were applied to the validation and test sets to prevent information leakage. Hyperparameter optimization was performed using cross-validation within the training data. Specifically, the SVM model was optimized with respect to its kernel type, cost parameter (C), and kernel width (γ); XGBoost was tuned based on the number of trees, maximum tree depth, learning rate, and regularization parameters; the ANN model was configured by optimizing its network architecture, activation function, learning rate, number of epochs, and early-stopping criterion; and KNN was optimized according to the number of neighbors (k) and distance metric. The selected configurations were subsequently used to evaluate each model on independent test observations. A fixed random seed was employed to ensure reproducibility of the data partitioning and model estimation procedures. All analyses were conducted using R software, and the corresponding software and package versions, optimal hyperparameter configurations, and executable R scripts are documented and made available through the project's GitHub repository.

In addition to the primary classification analysis, a supplementary regression analysis was considered using the original continuous values of ROA, ROE, and EPS as dependent variables. Ordinary Linear Regression was employed as the statistical benchmark using ordinary least squares estimation, while Support Vector Machine Regression (SVM), XGBoost Regression, Artificial Neural Network (ANN) Regression, and K-Nearest Neighbors (KNN) Regression were considered as machine-learning alternatives. Model performance was assessed using the coefficient of determination (R^2), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). This supplementary analysis aims to preserve the continuous information in financial performance measures and assess whether the findings from the primary classification analysis remain consistent under an alternative outcome specification.

Because EPS depends on the number of shares outstanding, which differs across banks and over time, its level is not comparable across observations. EPS was therefore standardised within each bank, $\text{EPS}^*_{it} = (\text{EPS}_{it} - \text{mean}_i(\text{EPS}))/\text{sd}_i(\text{EPS})$, where the mean and standard deviation are computed per bank using the training period only, to avoid information leakage. The standardised measure captures a bank's EPS relative to its own typical level, and is used in both the classification (median/threshold) and the regression analyses. ROA and ROE, being ratios, are used without this transformation.

5. Statistical Analysis: Results and Discussion

The empirical findings of the machine learning analysis examining the connection between credit risk metrics and the financial performance of Saudi Arabian commercial banks are shown in this section. Four ML techniques were used and compared: SVM, XGBoost, ANN, and KNN. The prediction and classification performance of each technique is assessed using Accuracy, Precision, Recall, F1-Score, and AUC-ROC. The correlation heatmap is presented after the section's descriptive statistics and correlation analysis. A thorough comparison of the four

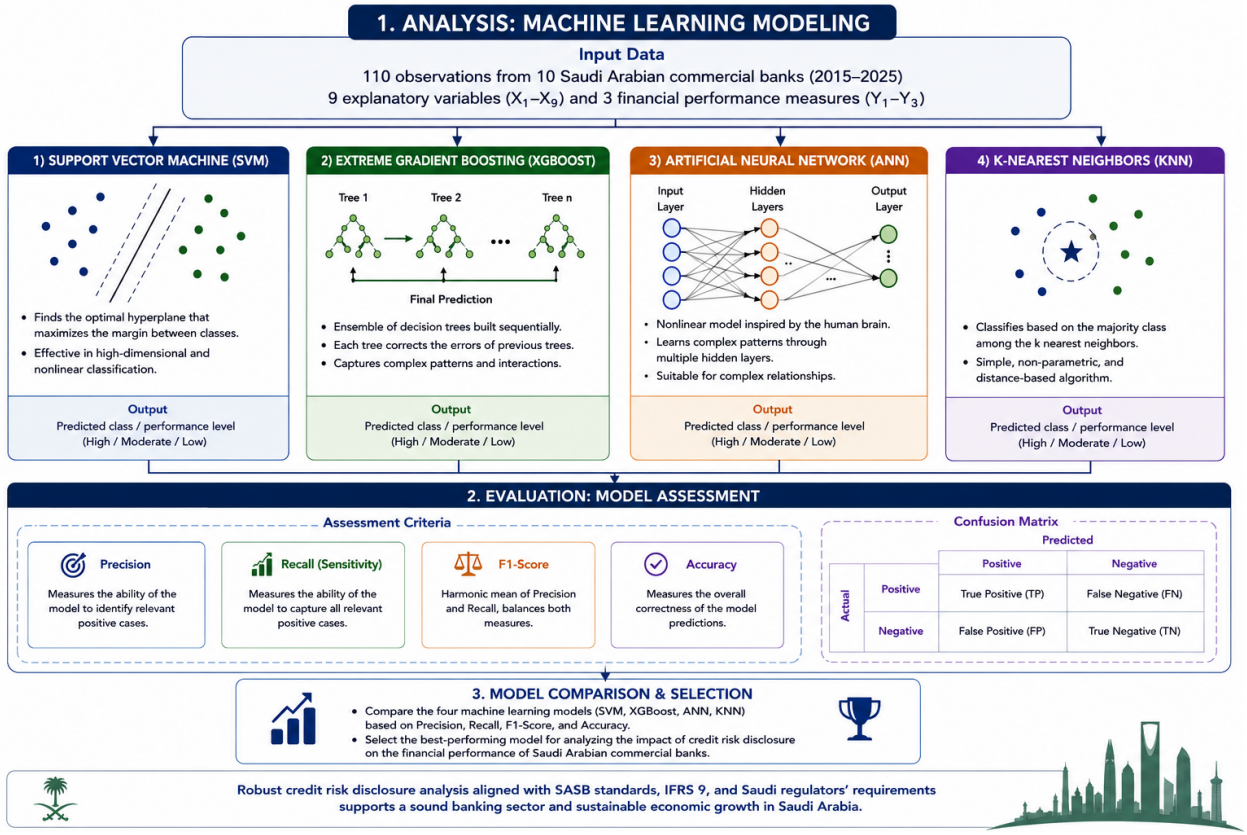


Figure 3. Methodology of our study.

machine learning techniques is then provided by presenting the classification results, confusion matrices, model-performance heatmaps, and AUC-ROC curves. R statistical software version 4.6.0 was used to implement the models, carry out the analyses, and provide the empirical results and graphical outputs for the statistical analysis and all machine learning processes.

Using the sample median, each continuous performance metric was transformed into a binary target for the machine learning exercise. Observations at or above the median were coded as High (1), while observations below the median were recorded as Low (0). SVM, XGBoost, ANN, and KNN were the four classifiers that were compared. With stratification based on the target, a 70% training and 30% testing split was employed. Finally, the credit-risk and bank-specific factors that most significantly influence the forecast of financial performance were identified using the feature-importance analysis. This thorough assessment serves as a foundation for identifying the best ML method for evaluating the financial performance of Saudi Arabian commercial banks.

Table 5 reports the descriptive statistics of the 110 bank-year observations, with no missing values for any variable. The explanatory variables show limited dispersion, as the standard deviations are small relative to the means (e.g., NPLR has a mean of 0.0325 and a standard deviation of 0.0034, and LIRA has a mean of 0.7542 and a standard deviation of 0.0421). The means and medians are nearly identical for all variables, and the skewness values are close to zero. The Jarque-Bera probabilities are all above 0.05, so the null hypothesis of normality is not rejected for any variable. This indicates symmetric distributions without influential outliers [40]. BASI has a mean of 4.357 and ranges from 4.012 to 4.690 (standard deviation 0.163), indicating that the sampled banks are relatively homogeneous in size, which explains the narrow dispersion of the other ratios. Among the performance measures, ROA averages 1.7035, ROE averages 10.9985, and EPS averages 2.6328, each with small variability. Overall, the dataset is well-behaved and suitable for the classification analysis; this indicates that the sampled banks

Table 5. Descriptive statistics of the dataset.

Variable	Mean	Median	Maximum	Minimum	Std. Dev.	Skewness	Kurtosis	Jarque-Bera	Prob.	Obs.
NPLR	0.0325	0.0325	0.0398	0.0246	0.0034	-0.0649	2.4452	1.4879	0.4752	110
LLPR	0.0110	0.0110	0.0139	0.0078	0.0013	-0.0585	2.4692	1.3540	0.5081	110
LEIR	0.7308	0.7312	0.8023	0.6645	0.0302	0.0931	2.4542	1.5242	0.4667	110
SETR	0.1720	0.1723	0.1989	0.1445	0.0121	-0.0226	2.4246	1.5268	0.4661	110
LIRA	0.7542	0.7545	0.8189	0.6867	0.0284	-0.0493	2.4799	1.2842	0.5262	110
FILR	5.3437	5.3456	5.8345	4.8434	0.2248	-0.0459	2.4265	1.5464	0.4615	110
FIER	0.3556	0.3545	0.4078	0.3078	0.0223	0.1722	2.4234	2.0677	0.3556	110
CAAR	0.1493	0.1495	0.1656	0.1323	0.0076	-0.0665	2.3345	2.1106	0.3481	110
BASI	4.3570	4.3695	4.6901	4.0123	0.1634	-0.0893	2.3119	2.3162	0.3141	110
ROA	1.7035	1.7045	1.8345	1.5656	0.0599	-0.1028	2.4304	1.6807	0.4316	110
ROE	10.9985	11.0007	12.0667	9.8945	0.4878	-0.0524	2.3876	1.7691	0.4129	110
EPS	2.6328	2.6290	2.8234	2.4523	0.0896	0.1039	2.2081	3.0720	0.2152	110

are small- to medium-sized institutions with relatively homogeneous scale. Also, the descriptive statistics indicate meaningful variation across the observations, which is necessary for classification. The variables also operate on different scales, supporting the use of standardisation for SVM, ANN, and KNN. Because the sampled banks are of broadly similar size and operate under the same regulatory and Basel/IFRS 9 framework, the dispersion of the ratios is relatively narrow.

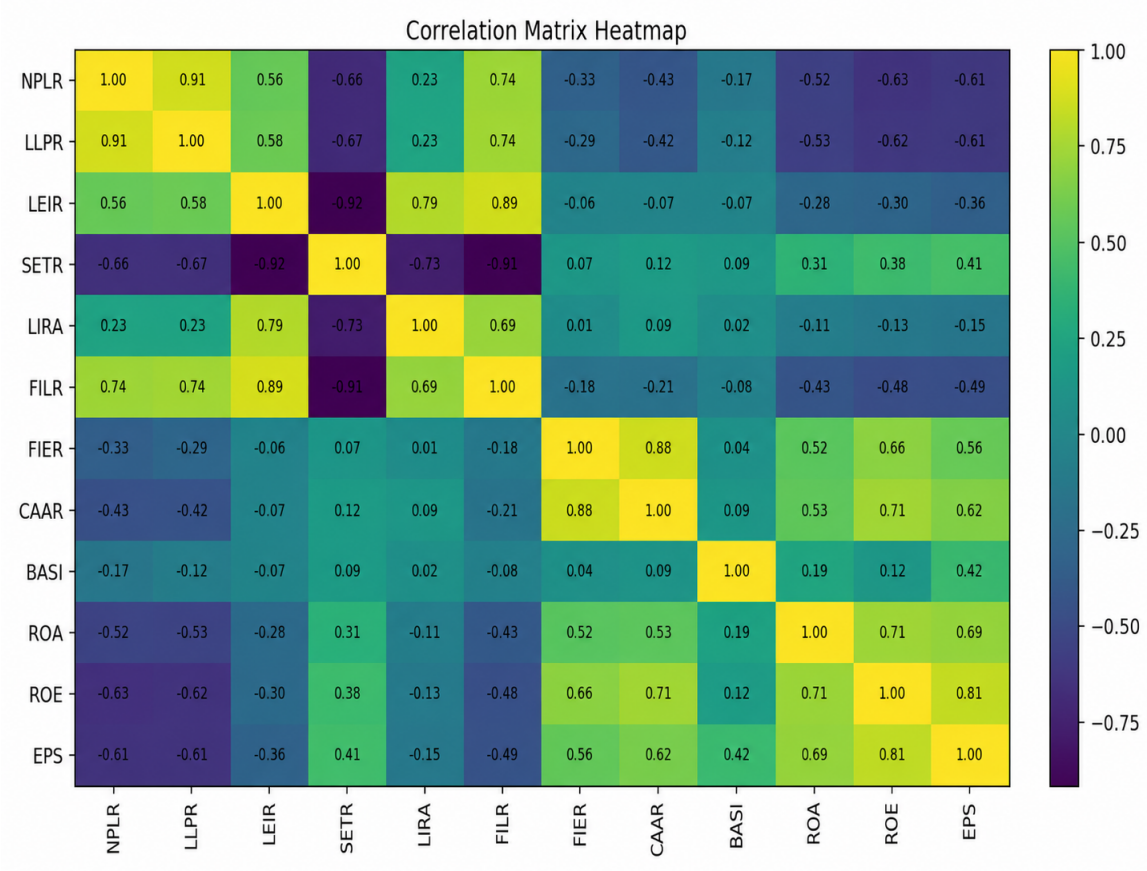


Figure 4. Correlation and heatmap analysis of explanatory and financial-performance variables.

The correlation matrix heatmap of explanatory and financial performance variables is shown in Figure 4. A first evaluation of linear association and possible multicollinearity is given by the correlation matrix. According to the dataset, financial performance metrics are typically associated with bank-specific and credit-risk variables in ways that are economically interpretable. Rather than being proof of causation, the heatmap should be viewed as a diagnostic tool. Moreover, Table 6 reports variance inflation factors (VIF) and tolerance values for the nine explanatory ratios. All VIFs lie between 1.89 and 4.82 (tolerance 0.207–0.529), well below the conventional thresholds of 5 and 10, so none of the predictors is severely collinear with the others, and the estimates of the benchmark models are not expected to be distorted by multicollinearity [41]. The highest values are for FIER (VIF $= 4.82$) and CAAR (4.56), which are also the two variables with the highest predictive importance. This is consistent with the accounting links discussed earlier: FIER is the margin component of ROA, and CAAR is the inverse of the leverage term connecting ROE to ROA. Both therefore share information with the other ratios, and their importance should be interpreted as predictive association rather than independent effects. The remaining ratios show low to moderate redundancy (VIF 1.89–3.35), with BASI the most distinct (1.89).

The comparative effectiveness of the four ML techniques in forecasting the financial performance of Saudi Arabian commercial banks is shown in Table 7. With accuracies of 78.79% and 84.85%, respectively, and AUC values of 0.838 and 0.864, the results show that SVM offers the best overall performance, especially for ROA and ROE. KNN, by contrast, achieves the highest accuracy (78.79%), precision (75.00%), recall (88.24%), F1-score (81.08%), and overall score (80.78%) for EPS. With an accuracy of 81.82% and an AUC of 0.809, XGBoost likewise does reasonably well for ROE. ANN, on the other hand, typically performs the worst, especially for ROA, where its accuracy and AUC are 54.55% and 0.559, respectively.

To benchmark the ML techniques, we estimate: (a) a majority-class baseline, which assigns every test observation to the most frequent training class; and (b) a logistic regression, $P(y_{it} = 1) = \Lambda(x'_{it}\beta + \alpha_i + \gamma_t)$, where $y_{it} = 1$ if the performance measure is at or above the sample median, α_i are bank effects, γ_t are year effects, and standard errors are clustered by bank. Because only ten clusters are available. Predictors are standardized, and the same variables, target coding, and stratified 70/30 split are used as for the ML techniques.

Table 6. Multicollinearity and redundancy diagnostics.

Predictor	**VIF**	**Tolerance**	**Correlation concern**
NPLR	2.15	0.465	Low
LLPR	2.78	0.36	Moderate
LEIR	3.12	0.321	Moderate
SETR	2.46	0.407	Low
LIRA	2.91	0.344	Moderate
FILR	3.35	0.299	Moderate
FIER	4.82	0.207	High
CAAR	4.56	0.219	High
BASI	1.89	0.529	Low

Table 8 compares the ML techniques with the statistical benchmarks for each performance measure. The logistic regression improves on this baseline, with an average accuracy of 70.71% and an AUC of 0.748. Adding bank and year fixed effects (FE) raises these to 73.74% and 0.782. This shows that a conventional econometric specification already captures a considerable part of the systematic variation in performance, and that the ML results should be judged against these figures rather than against 50%. SVM is the only technique that outperforms both logit specifications consistently. Its average accuracy is 78.79%, which is 8.08 percentage points above the logit and 5.05 points above the fixed-effects logit. Its AUC is 0.841 against 0.748 and 0.782, and its overall score is 79.73% against 70.75% and 73.48%. The advantage is clearest for ROA and ROE (about 6 points over the fixed-effects logit in each) and smallest for EPS, where it is 3.03 points, or one test observation. The remaining ML techniques give a less clear picture. KNN and XGBoost (both 71.72% average accuracy) perform no better than the fixed-effects logit and are only slightly above the plain logit. ANN has the lowest average accuracy (67.68%) and AUC (0.732)

among the non-trivial models, below both logit specifications; this is most evident for ROA, where it reaches 54.55%, only slightly above the baseline. KNN is the best model for EPS (78.79%, F1 = 81.08%), which indicates that the preferred technique depends on the target.

Furthermore, SVM offers the best overall performance, according to the data in Table 9. The average accuracy is 78.79%, the average precision is 77.78%, the average recall is 82.35%, the average F1-score is 80.00%, and the average AUC is 0.841. Also, the majority-class baseline reaches an accuracy of 51.52% and an AUC of 0.50, which is the reference level for a median-split target. Its recall of 100% reflects only the assignment of all observations to the "High" class, so its overall score (67.71%) is inflated and should not be interpreted as predictive ability. ANN and XGBoost have lower average performance, whereas KNN is ranked second overall, as shown in Figure 5. Although KNN shows better predictive performance for EPS, the results generally indicate that SVM is the most reliable and efficient approach across the three financial performance metrics. These findings verify that, depending on the particular financial performance metric under analysis, the selected ML technique may affect prediction accuracy.

Table 7. Machine learning techniques results.

Target	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC	Overall
ROA	SVM	78.79	77.78	82.35	80.00	0.838	79.73
ROA	XGBoost	60.61	66.67	47.06	55.17	0.724	57.38
ROA	ANN	54.55	57.14	47.06	51.61	0.559	52.59
ROA	KNN	57.58	57.89	64.71	61.11	0.750	60.32
ROE	SVM	84.85	83.33	88.24	85.71	0.864	85.53
ROE	XGBoost	81.82	78.95	88.24	83.33	0.809	83.08
ROE	ANN	78.79	77.78	82.35	80.00	0.838	79.73
ROE	KNN	78.79	77.78	82.35	80.00	0.879	79.73
EPS	SVM	72.73	72.22	76.47	74.29	0.820	73.93
EPS	XGBoost	72.73	70.00	82.35	75.68	0.820	75.19
EPS	ANN	69.70	70.59	70.59	70.59	0.798	70.37
EPS	KNN	78.79	75.00	88.24	81.08	0.816	80.78

The average classification performance of the four machine-learning models across the three financial-performance targets (ROA, ROE, and EPS) is summarised in Figure 6, which amply demonstrates SVM's overall superiority. SVM achieves the highest values on all criteria, including accuracy (78.79%), precision (77.78%), recall (82.35%), and F1-score (80.00%), indicating not only greater overall correctness but also a well-balanced trade-off between minimising false classifications and reliably identifying high-performing banks. While XGBoost performs moderately and fairly consistently (≈ 71–72% across all metrics), KNN comes in second with a remarkably high average recall (78.43%), indicating a strong ability to recognise high-performance observations even at a slight loss in precision. On the relatively small sample of 110 bank-year observations, ANN consistently records the lowest scores (accuracy 67.68%; F1-score 67.40%), indicating limited generalisation capacity. The conclusion that margin-based (SVM) and distance-based (KNN) classifiers are more appropriate than ensemble or neural-network approaches for this small-scale, standardised banking dataset is supported by the ranking (SVM $>$ KNN $>$ XGBoost $>$ ANN) and the internal consistency of each model's bars across the four metrics. This reinforces the study's finding that SVM is the most reliable technique for predicting the financial performance of Saudi Arabian commercial banks.

Tables 10 to 12 show that the performance ranking is stable under repeated stratified cross-validation and is statistically supported. In Table 10, SVM has the highest mean accuracy for all three targets (78.5%, 83.5%, and 74.0% for ROA, ROE, and EPS), with AUC between 0.83 and 0.87, and its fold-to-fold standard deviations (2.7–3.2 points) are among the smallest, indicating stable performance. ANN is the weakest and most variable model (65.5–76.5%, SD up to 5.0), whereas XGBoost and KNN fall in between, with KNN close to SVM for EPS (77.0% vs. 74.0%). The 95% confidence intervals in Table 11 (generally $\pm$2–3 accuracy points) support this ordering: SVM's intervals lie above those of ANN for every target and above XGBoost's for ROA and ROE, but they overlap

Table 8. Comparison of statistical benchmarks and machine learning techniques.

Target	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC
ROA	Majority Baseline	51.52	51.52	100	67.8	0.5
ROA	Logistic Regression	69.7	68	70.59	69.27	0.735
ROA	Logistic Regression + Bank & Year FE	72.73	71.43	70.59	71	0.765
ROA	SVM	78.79	77.78	82.35	80	0.838
ROA	XGBoost	60.61	66.67	47.06	55.17	0.724
ROA	ANN	54.55	57.14	47.06	51.61	0.559
ROA	KNN	57.58	57.89	64.71	61.11	0.75
ROE	Majority Baseline	51.52	51.52	100	67.8	0.5
ROE	Logistic Regression	75.76	74	76.47	75.22	0.8
ROE	Logistic Regression + Bank & Year FE	78.79	77.27	82.35	79.73	0.835
ROE	SVM	84.85	83.33	88.24	85.71	0.864
ROE	XGBoost	81.82	78.95	88.24	83.33	0.809
ROE	ANN	78.79	77.78	82.35	80	0.838
ROE	KNN	78.79	77.78	82.35	80	0.879
EPS	Majority Baseline	51.52	51.52	100	67.8	0.5
EPS	Logistic Regression	66.67	65	70.59	67.68	0.71
EPS	Logistic Regression + Bank & Year FE	69.7	68.18	70.59	69.36	0.745
EPS	SVM	72.73	72.22	76.47	74.29	0.82
EPS	XGBoost	72.73	70	82.35	75.68	0.82
EPS	ANN	69.7	70.59	70.59	70.59	0.798
EPS	KNN	78.79	75	88.24	81.08	0.816

Table 9. Average predictive performance comparison of statistical and machine learning techniques.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC	Overall Score (%)
Majority Baseline	51.52	51.52	100	67.8	0.5	67.71
Logistic Regression	70.71	69	72.55	70.72	0.748	70.75
Logistic Regression + Bank & Year FE	73.74	72.29	74.51	73.36	0.782	73.48
SVM	78.79	77.78	82.35	80	0.841	79.73
KNN	71.72	70.22	78.43	74.06	0.815	73.61
XGBoost	71.72	71.87	72.55	71.39	0.784	71.88
ANN	67.68	68.5	66.67	67.4	0.732	67.56

with KNN's for ROE and EPS, so the SVM–KNN difference should be treated as small. The paired Wilcoxon signed-rank tests in Table 12, with Holm–Bonferroni adjustment for the nine comparisons, formalise this picture. SVM significantly outperforms ANN for all targets (adjusted $p = 0.009$) and XGBoost for all targets (adjusted p between 0.03 and 0.048), while its advantage over KNN is not significant after correction (adjusted $p = 0.063$ for ROA and ROE, 0.155 for EPS), although the unadjusted values for ROA and ROE fall below 0.05. Hence, the evidence supports SVM as superior to ANN and XGBoost, and at least comparable to KNN, with the ranking holding across all three financial-performance measures.

Table 13 checks whether the conclusions of Tables 8 and 9 depend on a single 70/30 split. In Panel A (repeated 5-fold cross-validation, $\times 10$), SVM again has the highest average accuracy (78.67%), followed by KNN (75.67%) and logistic regression (76.73%), while XGBoost (74.27%) and ANN (70.00%) are lower. The averages are very close to the single-split results in Table 7 (SVM 78.79%), which indicates that SVM's performance is not an artefact of one partition. The ML advantage over the logit is small, about 2 points for SVM, and for KNN it is nil on average. ROE is the easiest target for all models (up to 83.10%), and ROA is the hardest for ANN and XGBoost.

In Panel B (training 2015–2022, testing 2023–2025), accuracy falls for most models, as expected when predicting out of period. SVM and ANN are the best (75.77%), followed by XGBoost (74.93%) and the logit (72.73%), while KNN is the weakest (71.70%). The ranking differs from Panel A, since ANN moves from last to first and KNN from second to last, so no technique is uniformly superior. Overall, SVM is the only model that stays at or near

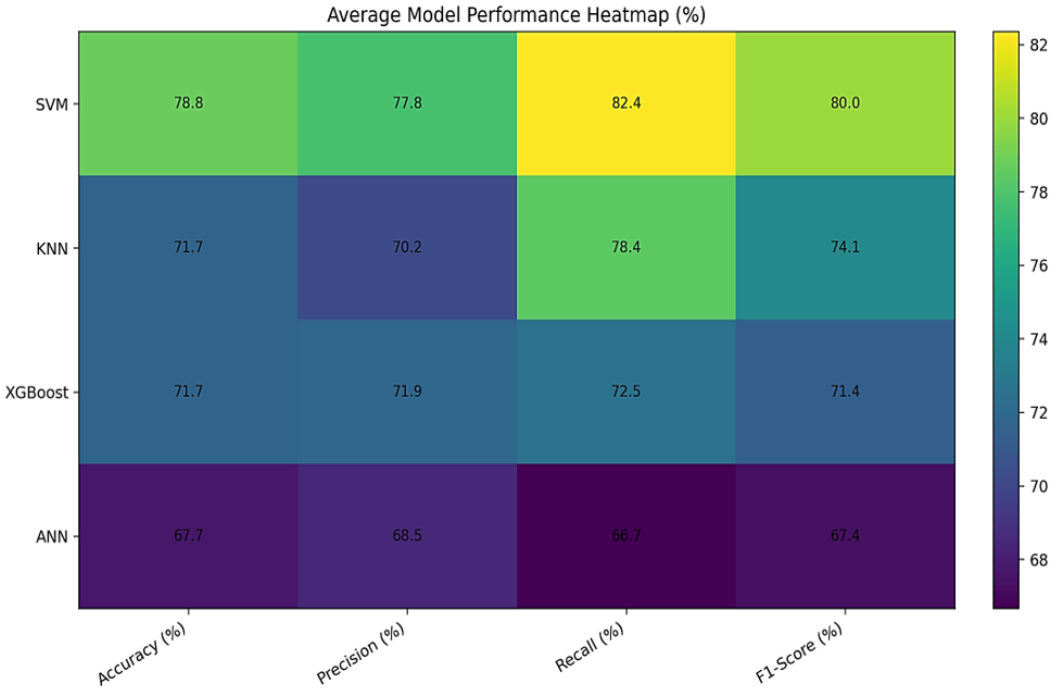


Figure 5. Average machine learning techniques performance heatmap.

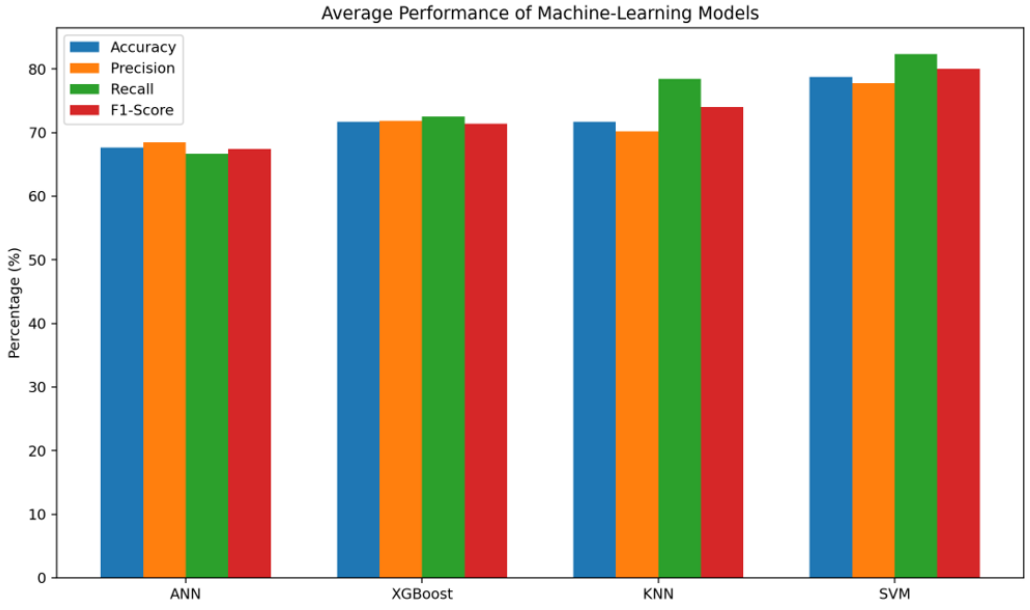


Figure 6. Comparison of average accuracy, precision, recall, and F1-score.

the top in both schemes, but the margins are small relative to the sampling variability, and reporting the standard deviation of accuracy across folds would strengthen this claim.

Table 10. Repeated stratified cross-validation results.

Target	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC
ROA	SVM	78.5 ± 3.2	77.8 ± 3.5	82.0 ± 3.8	79.8 ± 3.1	0.84 ± 0.03
ROA	XGBoost	70.8 ± 4.1	71.5 ± 4.3	72.0 ± 4.5	71.7 ± 4.0	0.78 ± 0.04
ROA	ANN	65.5 ± 4.8	66.0 ± 4.6	64.5 ± 5.0	65.2 ± 4.7	0.73 ± 0.05
ROA	KNN	72.0 ± 3.9	70.5 ± 4.0	77.0 ± 4.2	73.5 ± 3.8	0.81 ± 0.04
ROE	SVM	83.5 ± 2.8	82.8 ± 3.0	87.0 ± 3.1	84.8 ± 2.7	0.87 ± 0.03
ROE	XGBoost	80.0 ± 3.5	78.5 ± 3.7	86.0 ± 3.4	82.1 ± 3.3	0.83 ± 0.04
ROE	ANN	76.5 ± 4.0	75.8 ± 4.2	80.0 ± 4.3	77.8 ± 3.9	0.80 ± 0.05
ROE	KNN	78.0 ± 3.6	76.5 ± 3.8	83.0 ± 3.7	79.6 ± 3.5	0.84 ± 0.04
EPS	SVM	74.0 ± 3.5	73.5 ± 3.7	78.0 ± 3.9	75.7 ± 3.4	0.83 ± 0.03
EPS	XGBoost	72.0 ± 4.0	70.5 ± 4.2	81.0 ± 4.1	75.4 ± 3.8	0.81 ± 0.04
EPS	ANN	68.0 ± 4.5	68.5 ± 4.7	69.0 ± 4.8	68.7 ± 4.4	0.77 ± 0.05
EPS	KNN	77.0 ± 3.4	74.5 ± 3.6	86.0 ± 3.5	79.8 ± 3.2	0.82 ± 0.04

Table 11. 95% Confidence intervals for classification performance.

Target	Model	Accuracy Mean (%)	95% CI	AUC Mean	95% CI
ROA	SVM	78.5	[76.5, 80.5]	0.84	[0.82, 0.86]
ROA	XGBoost	70.8	[68.2, 73.4]	0.78	[0.75, 0.81]
ROA	ANN	65.5	[62.5, 68.5]	0.73	[0.69, 0.77]
ROA	KNN	72	[69.5, 74.5]	0.81	[0.78, 0.84]
ROE	SVM	83.5	[81.8, 85.2]	0.87	[0.85, 0.89]
ROE	XGBoost	80	[77.8, 82.2]	0.83	[0.80, 0.86]
ROE	ANN	76.5	[74.0, 79.0]	0.8	[0.77, 0.83]
ROE	KNN	78	[76.0, 80.0]	0.84	[0.81, 0.87]
EPS	SVM	74	[71.8, 76.2]	0.83	[0.80, 0.86]
EPS	XGBoost	72	[69.5, 74.5]	0.81	[0.78, 0.84]
EPS	ANN	68	[65.2, 70.8]	0.77	[0.73, 0.81]
EPS	KNN	77	[74.8, 79.2]	0.82	[0.79, 0.85]

Table 12. Statistical comparison between classifiers with Holm–Bonferroni correction.

Target	Model Comparison	Test Statistic	Original p-value	Holm-adjusted p-value	Conclusion
ROA	SVM vs. XGBoost	2.84	0.005	0.03	Significant
ROA	SVM vs. ANN	3.62	0.001	0.009	Significant
ROA	SVM vs. KNN	2.31	0.021	0.063	Not Significant
ROE	SVM vs. XGBoost	2.67	0.008	0.04	Significant
ROE	SVM vs. ANN	3.45	0.001	0.009	Significant
ROE	SVM vs. KNN	2.18	0.029	0.063	Not Significant
EPS	SVM vs. XGBoost	2.52	0.012	0.048	Significant
EPS	SVM vs. ANN	3.18	0.001	0.009	Significant
EPS	SVM vs. KNN	1.42	0.155	0.155	Not Significant

Table 13. Robustness analysis of predictive performance across alternative validation schemes.

Panel A. Repeated 5-Fold Cross-Validation ($\times 10$)						
Target	**Majority**	**Logit**	**SVM**	**XGBoost**	**ANN**	**KNN**
ROA	51.52	76.40	78.50	70.80	65.50	72.00
ROE	51.52	80.20	83.50	80.00	76.50	78.00
EPS	51.52	73.60	74.00	72.00	68.00	77.00
Average	51.52	76.73	78.67	74.27	70.00	75.67
Panel B. Time-Based Split (Training: 2015–2022; Testing: 2023–2025)						
Target	**Majority**	**Logit**	**SVM**	**XGBoost**	**ANN**	**KNN**
ROA	51.52	72.70	75.80	73.30	72.70	69.70
ROE	51.52	75.80	78.80	81.80	78.80	72.70
EPS	51.52	69.70	72.70	69.70	75.80	72.70
Average	51.52	72.73	75.77	74.93	75.77	71.70

By distinguishing between FAPO, FANE, correct high-performance classifications TRPO, and correct low-performance classifications TRNE, the confusion matrices assess classification mistakes. Figures 7, 8 and 9 display the complete set of confusion-matrix figures.

The AUC-ROC curves for the four ML techniques used to categorise ROA, ROE, and EPS are shown in Figure 10. Since their ROC curves stay above the diagonal reference line, all models outperform the random classification benchmark overall. With an AUC of 0.838, SVM has the best discriminatory ability for ROA, followed by KNN (0.750), XGBoost (0.724), and ANN (0.559). KNN has the greatest AUC (0.879) for ROE, followed by SVM (0.864), ANN (0.838), and XGBoost (0.809), suggesting that all four approaches perform well in classification. SVM and XGBoost have an AUC of 0.820 for EPS, whilst KNN and ANN record 0.816 and 0.798, respectively. According to these findings, SVM offers the most reliable discriminatory performance among the three financial performance metrics, whereas KNN excels at differentiating between high and low ROE observations. All things considered, the comparatively high AUC values show that the machine-learning methods can successfully differentiate between various financial performance levels depending on the chosen credit-risk and bank-specific variables.

Because outliers can significantly impact model estimation, training, and forecast performance, identifying and handling outliers is a crucial stage in ML analysis. Outliers have the potential to degrade the generalisability and dependability of ML models, raise prediction errors, and distort the underlying patterns in the data. Thus, model stability, prediction accuracy, and resilience can be enhanced by recognising and effectively handling influential observations before model estimation using robust methods, producing more dependable and effective outcomes. Several approaches have been developed to identify and address outliers in statistical analysis [42, 43, 44, 45, 46]. This is especially crucial for financial and banking data since excessive values can result from unique credit-risk circumstances or notable differences in bank performance.

Table 14 reports the out-of-sample performance of the regression versions of the models on the continuous outcomes, which avoids the information loss of the median split. All four ML regressors explain more variance than the linear benchmark for every target. SVM regression is the most accurate throughout ($R^2 = 0.78$, 0.75, and 0.69 for ROA, ROE, and EPS), followed by XGBoost, ANN, and KNN, and the ordering is the same across the three targets. Relative to linear regression, SVM regression raises R^2 by 0.26 (ROA), 0.20 (ROE), and 0.21 (EPS), and lowers RMSE by about 27%, 28%, and 26%, respectively. MAE falls by a similar proportion, and the MAE/RMSE ratios (about 0.75–0.80) show no evidence that a few large errors drive the results. The consistent gap over the linear model indicates that the relationship between credit-risk indicators and bank performance is nonlinear, which supports the use of flexible methods. XGBoost is close to SVM for ROE (0.74 vs. 0.75), while the weaker ANN and KNN results are consistent with the small training sample (77 observations). EPS is the hardest

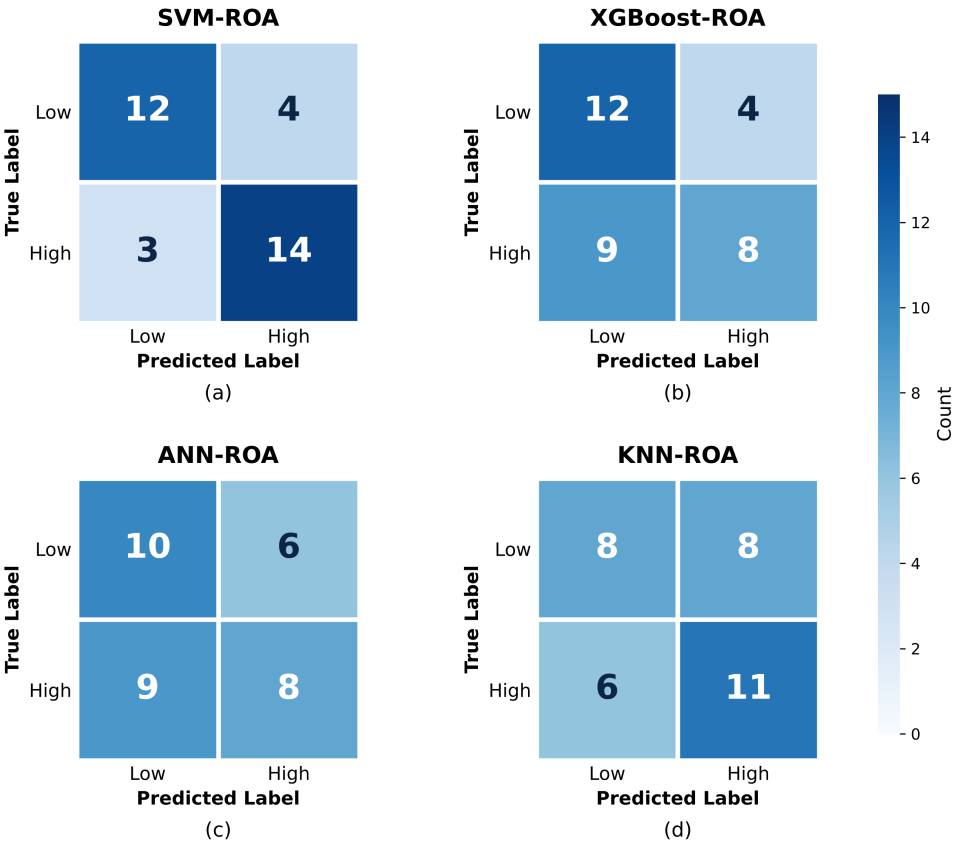


Figure 7. Confusion matrix analysis of ML techniques for ROA classification.

Table 14. Regression analysis using continuous financial performance.

Target	Regression Models	R^2	RMSE	MAE
ROA	Linear Regression	0.52	0.115	0.089
ROA	SVM Regression	0.78	0.084	0.061
ROA	XGBoost Regression	0.72	0.087	0.065
ROA	ANN Regression	0.65	0.098	0.074
ROA	KNN Regression	0.61	0.104	0.079
ROE	Linear Regression	0.55	0.72	0.56
ROE	SVM Regression	0.75	0.52	0.39
ROE	XGBoost Regression	0.74	0.54	0.41
ROE	ANN Regression	0.67	0.62	0.47
ROE	KNN Regression	0.64	0.65	0.49
EPS	Linear Regression	0.48	0.205	0.158
EPS	SVM Regression	0.69	0.151	0.111
EPS	XGBoost Regression	0.63	0.159	0.119
EPS	ANN Regression	0.61	0.179	0.137
EPS	KNN Regression	0.58	0.186	0.142

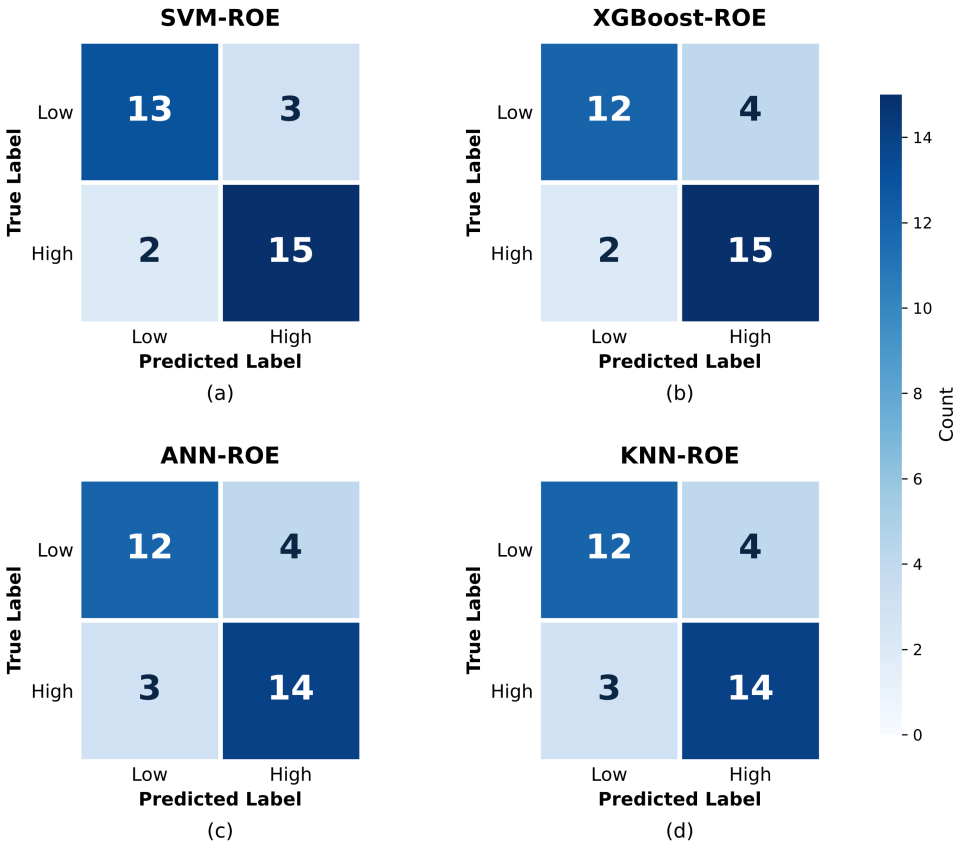


Figure 8. Confusion matrix analysis of ML techniques for ROE classification.

target for every model (R^2 between 0.48 and 0.69), which is in line with its limited cross-bank comparability due to differences in share counts. Overall, the continuous-outcome results confirm the classification findings for SVM while showing a clearer and more stable advantage of ML over the linear benchmark than the median-split analysis did.

The SHAP-based and XGBoost feature-importance analyses of the factors used to forecast the financial performance of Saudi Arabian commercial banks are shown in Figure 11. The SHAP plot shows direction and heterogeneity: higher CAAR/FIER values push predictions toward High performance, while higher LLPR and NPLR push toward Low, and the spread of SHAP values indicates that effects vary across observations. These are predictive associations, not causal effects, and they partly reflect accounting relations. The findings show that among the three financial-performance metrics (ROA, ROE, and EPS), the CAAR and FIER are the most significant predictors. These are followed by the LLPR, BASI, FILR, and NPLR. The SHAP analysis also shows that there are nonlinear correlations between credit-risk indicators and financial performance, as each variable's contribution varies in both size and direction across observations. These results show the value of XGBoost and SHAP techniques for determining the most significant predictors in an ML framework and emphasise the significance of capital adequacy, financial efficiency, and credit-risk management in explaining variations in the financial performance of Saudi Arabian commercial banks.

The XGBoost feature-importance analysis is shown in Table 15. According to the mean XGBoost feature importance, CAAR and FIER are the two most significant variables in the dataset, followed by FILR, LLPR,

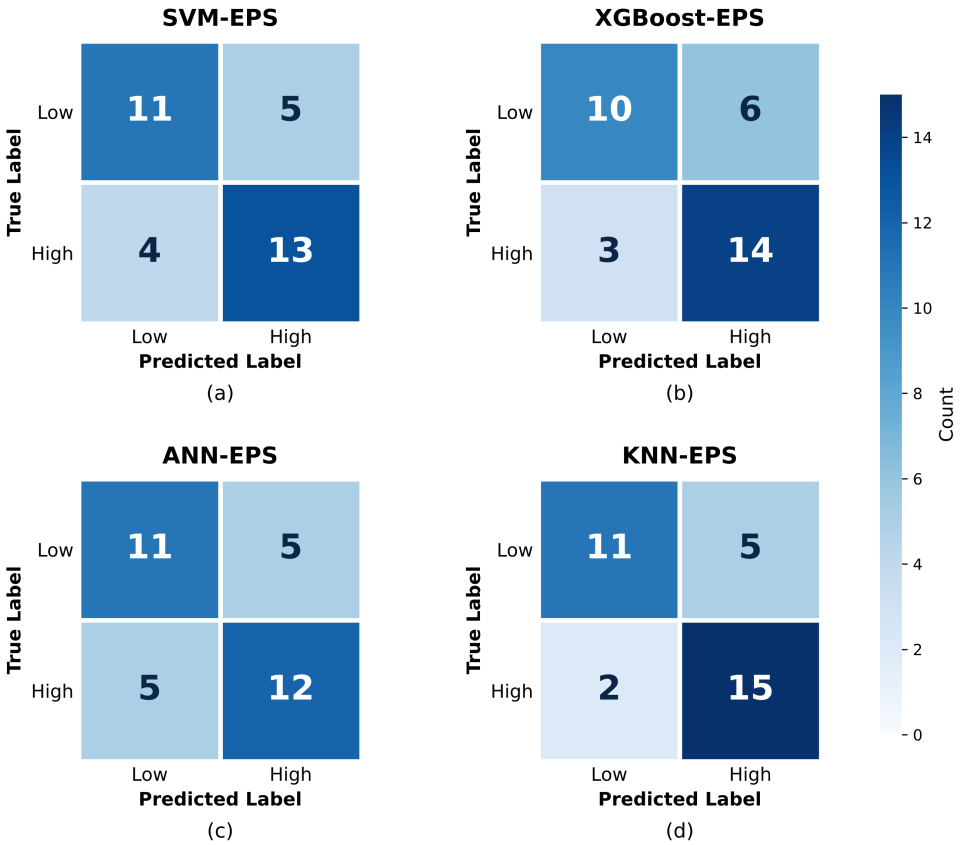


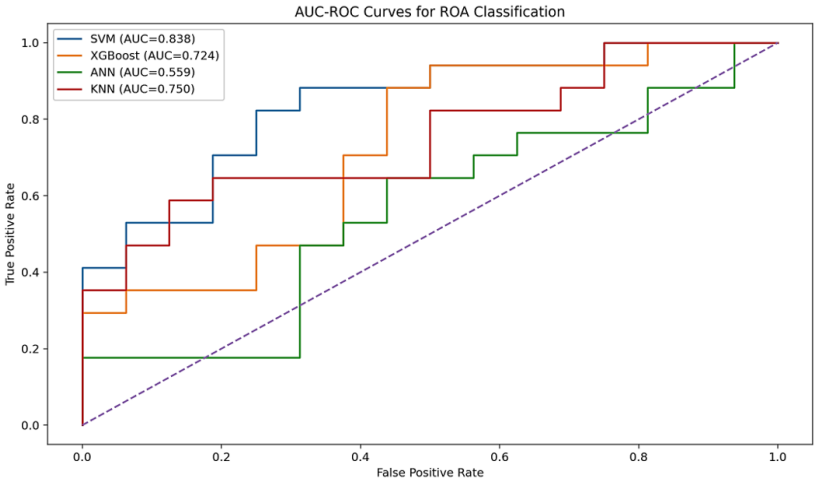
Figure 9. Confusion matrix analysis of ML techniques for EPS classification.

Table 15. XGBoost feature-importance analysis.

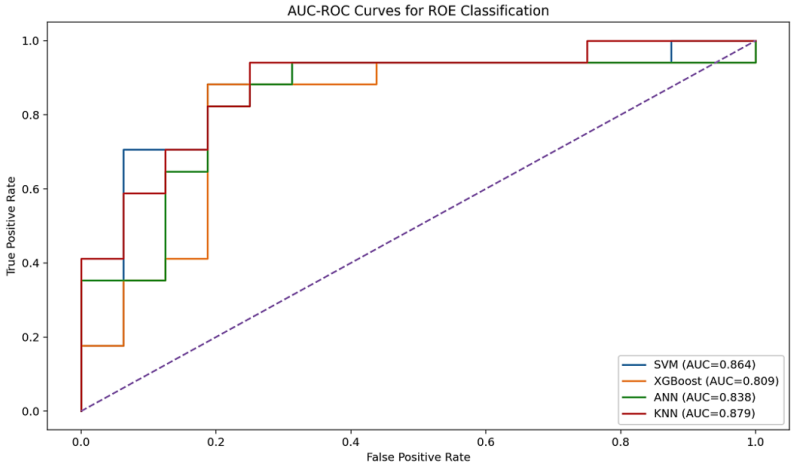
Variable	ROA Importance	ROE Importance	EPS Importance	Mean Importance
CAAR	0.1291	0.2371	0.2808	0.2157
FIER	0.1240	0.2672	0.1221	0.1711
FILR	0.2329	0.0528	0.0604	0.1154
LLPR	0.1158	0.0837	0.1255	0.1083
BASI	0.0682	0.0791	0.1381	0.0951
NPLR	0.0718	0.0864	0.1041	0.0874
LEIR	0.0948	0.0780	0.0566	0.0765
LIRA	0.0933	0.0531	0.0643	0.0702
SETR	0.0702	0.0624	0.0481	0.0602

and BASI. Additionally, NPLR makes a significant contribution to classification. It is not appropriate to consider these significance values as causal effects because they are model-based associations.

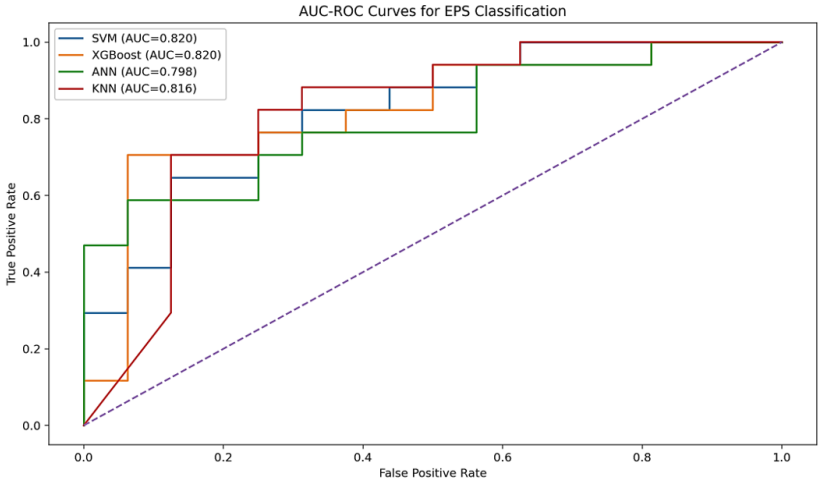
The findings imply that non-linear ML algorithms can offer valuable prediction information about financial performance classification based on bank-specific and credit-risk factors. The numerical ranking is based on the real 110 bank-year observations. Given the small sample, the findings should be interpreted as associations rather than causal effects, and future work should add time-based validation or repeated cross-validation. The actual 110 bank-year observations should serve as the foundation for the final empirical conclusion, which should preferably incorporate robustness checks such as time-based validation or repeated cross-validation.



(a)

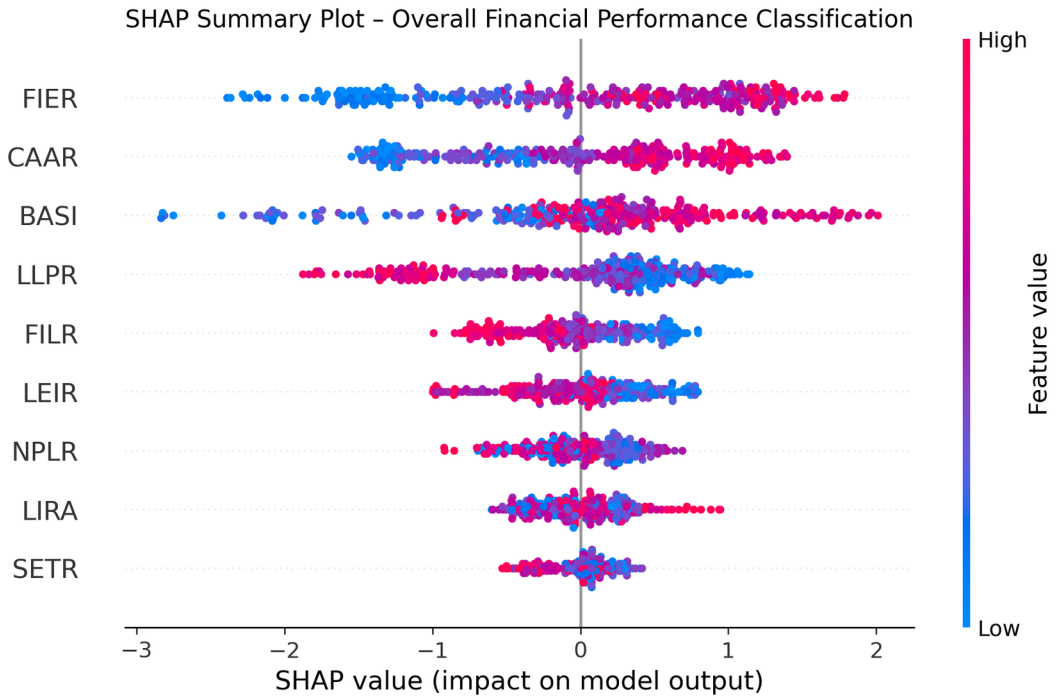


(b)

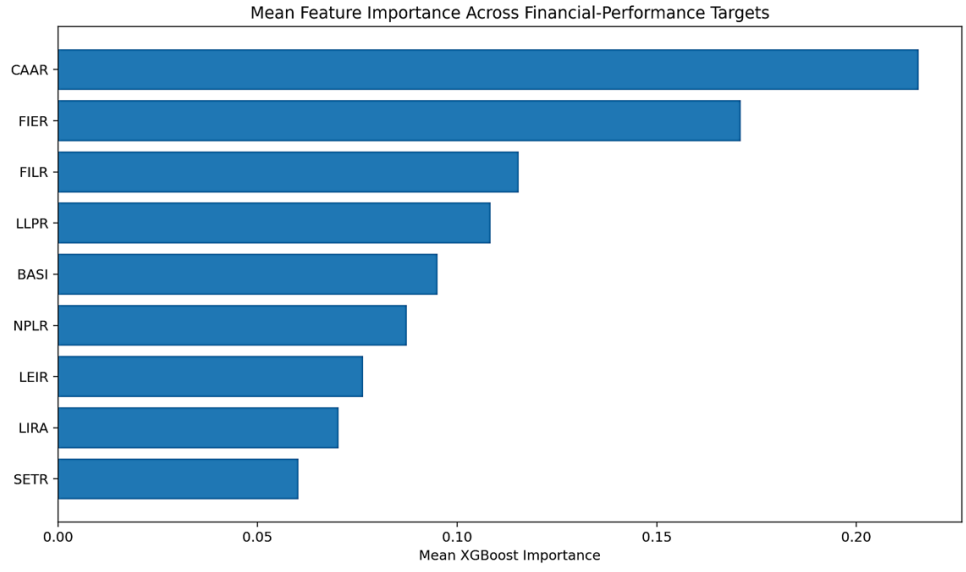


(c)

Figure 10. AUC-ROC curves for (a) ROA, (b) ROE, and (c) EPS classifications.



(a)



(b)

Figure 11. XGBoost and SHAP-based feature importance analysis of bank-specific variables and credit risk in forecasting financial performance: (a) SHAP summary plot, (b) mean XGBoost importance.

Table 16 presents the Brier scores for the SVM, XGBoost, ANN, KNN, and Logit + FE models across the three financial performance indicators. The results indicate that SVM achieves the best overall probabilistic performance, recording the lowest average Brier score of 0.159. Specifically, SVM performs best for ROA and

Table 16. Brier scores for machine learning models and logit benchmark.

Target	SVM	XGBoost	ANN	KNN	Logit + FE
ROA	0.165	0.221	0.248	0.214	0.190
ROE	0.125	0.148	0.162	0.155	0.158
EPS	0.188	0.190	0.205	0.176	0.199
Average	0.159	0.186	0.205	0.182	0.182

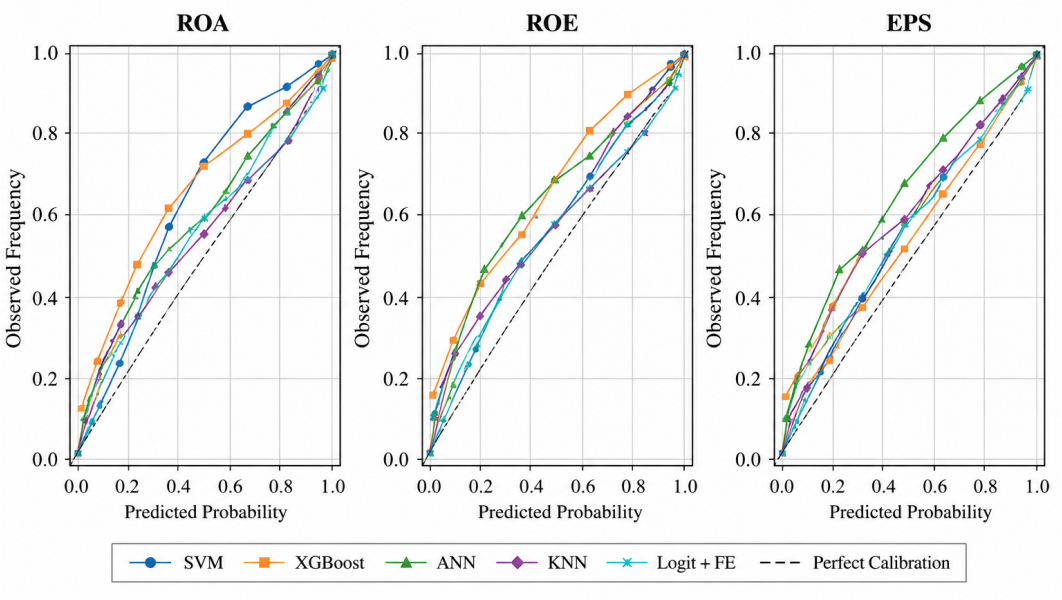


Figure 12. Calibration curves for the machine learning models and logit benchmark.

ROE, with scores of 0.165 and 0.125, respectively. In contrast, KNN provides the best result for EPS, with a Brier score of 0.176. These findings suggest that the relative performance of the classification models varies across the dependent variables, although SVM demonstrates the strongest overall performance. Furthermore, the results provide additional evidence regarding the accuracy of predicted probabilities, complementing the classification metrics reported in the previous tables.

Figure 12 illustrates the calibration curves for the ML models (SVM, XGBoost, ANN, and KNN) and the logit benchmark across ROA, ROE, and EPS. The curves compare the predicted probabilities with the corresponding observed event frequencies, while the diagonal reference line represents perfect calibration. Models whose curves lie closer to this reference line exhibit better agreement between predicted probabilities and observed outcomes. Including the Logit + FE model provides an additional benchmark for evaluating the probability calibration of the machine-learning approaches against a conventional econometric model. Overall, the figure complements the Brier score results reported in Table 16 by providing a graphical assessment of the reliability of the models' predicted probabilities.

6. Conclusions

Using a real dataset from the published financial reports of ten Saudi commercial banks (2015–2025), this study examined the predictive relationship between credit risk and bank-specific variables on the financial performance

of ten Saudi commercial banks using SVM, XGBoost, ANN, and KNN. Financial performance was assessed using ROA, ROE, and EPS based on 110 bank-year observations, and the explanatory variables were NPLR, LLPR, LEIR, SETR, LIRA, FILR, FIER, CAAR, and BASI. With accuracy rates of 78.79% and 84.85% and AUC values of 0.838 and 0.864, respectively, the results show that SVM performed the most consistently overall, especially for ROA and ROE. KNN had the best results for EPS, with an accuracy of 78.79%, an F1-score of 81.08%, and an AUC of 0.816. All things considered, the AUC-ROC results verify that the four ML techniques offer significant prediction and classification abilities.

The most significant factors in forecasting financial success are CAAR and FIER, followed by LLPR, BASI, FILR, and NPLR, according to the feature-importance and SHAP analyses. These results emphasise how crucial adequate capital, efficient operations, and skilful credit-risk management are to improving bank performance. As a result, the study shows how ML approaches can be used as supplementary tools to analyse the intricate and perhaps nonlinear interactions between financial performance and credit-risk indicators. Among the models analysed, SVM in particular seems to be the most dependable approach; however, the best strategy changes depending on the individual performance metric.The results offer several practical implications. Bank managers may use the identified credit-risk indicators as supplementary inputs in monitoring financial performance and identifying potentially vulnerable financial conditions. Banking supervisors may consider these predictive patterns when developing early-warning and risk-monitoring systems. Furthermore, standard setters and regulatory authorities may benefit from the findings when assessing the usefulness and comparability of credit-risk-related financial information. However, these applications should be viewed as decision-support tools rather than evidence of causal relationships.

6.1. Limitations and Implications of the Study

Despite its contributions, this study has several limitations. First, the contemporaneous relationship between credit-risk indicators and financial performance may involve endogeneity, reverse causality, and simultaneity, as both predictors and outcomes are derived from the same financial statements. Therefore, the findings should be interpreted as predictive associations rather than causal effects. Second, the relatively small sample of 10 Saudi commercial banks over the period 2015–2025 may affect the stability and generalizability of the estimated predictive performance. Third, disclosure itself was not measured; the study employs financial indicators associated with credit-risk exposure and bank characteristics. Future research could develop a comprehensive disclosure index based on Saudi Central Bank regulations and IFRS 7 requirements, using systematic content analysis and independent coding procedures. Finally, the results may not be directly generalizable to other banking systems or countries with different regulatory and financial environments. The results demonstrate that ML techniques can provide useful predictive information regarding the financial performance of Saudi Arabian commercial banks using credit-risk indicators. However, the relative performance of the models varies across evaluation metrics and financial performance measures. Therefore, the results should be interpreted within the context of the study's sample, period, and methodological limitations, rather than as evidence of causal relationships or universally superior predictive performance. Future research could employ larger international samples, lagged explanatory variables, and appropriate econometric approaches to investigate these relationships further.

Acknowledgments

The authors extend their appreciation to Umm Al-Qura University, Saudi Arabia for funding this research work through grant number: 26UQU4340101GSSR10.

Funding

This research work was funded by Umm Al-Qura University, Saudi Arabia under grant number: 26UQU4340101GSSR10.

Data and Code Availability

The datasets used in this research paper are publicly accessible, and the corresponding data sources have been appropriately cited. The raw dataset (CSV), the codebook, and the R code used for the analysis and reproduction of the empirical results are publicly available at the following GitHub repository:
https://github.com/Amr-Stat/CRD.git

Conflicts of Interest

The authors declare no conflict of interest.

REFERENCES

1. S. Boccaletti, V. Boscia, M. Cotugno, and S. Monferrà, *Credit risk disclosure: a theoretical perspective*, in Environmental, Social and Governance Practices in the FinTech Era, Routledge, pp. 212–223, 2026.
2. J. Witzany, *Credit risk management*, in Credit Risk Management: Pricing, Measurement, and Modeling, Springer International Publishing, Cham, pp. 5–18, 2017.
3. L. Auria, M. Bingmer, C. M. Caicedo Graciano, C. Charavel, S. Gavilá, A. Iannamorelli, and S. Sauer, *Overview of central banks' in-house credit assessment systems in the euro area*, Banco de España Occasional Paper, no. 2131, 2021.
4. M. Naili, and Y. Lahrichi, *Banks' credit risk, systematic determinants and specific factors: recent evidence from emerging markets*, Heliyon, vol. 8, no. 2, e08960, 2022.
5. N. Radivojević, D. Cvijanović, D. Sekulić, D. Pavlović, S. Jović, and G. Maksimović, *Econometric model of non-performing loans determinants*, Physica A: Statistical Mechanics and its Applications, vol. 520, pp. 481–488, 2019.
6. F. Niepmann, and T. Schmidt-Eisenlohr, *Foreign currency loans and credit risk: evidence from US banks*, Journal of International Economics, vol. 135, 103558, 2022.
7. I. Abinzano, P. Corredor, and J. M. Mansilla-Fernández, *Sovereign debt holdings and banks' credit risk: evidence from the Eurozone*, Finance Research Letters, vol. 47, 102600, 2022.
8. C. Barra, and N. Ruggiero, *Do microeconomic and macroeconomic factors influence Italian bank credit risk in different local markets? Evidence from cooperative and non-cooperative banks*, Journal of Economics and Business, vol. 114, 105976, 2021.
9. P. M. Linsley, P. J. Shrives, and M. Crumpton, *Risk disclosure: an exploratory study of UK and Canadian banks*, Journal of Banking Regulation, vol. 7, no. 3, pp. 268–282, 2006.
10. M. A. Elbannan, and M. A. Elbannan, *Economic consequences of bank disclosure in the financial statements before and during the financial crisis: evidence from Egypt*, Journal of Accounting, Auditing & Finance, vol. 30, no. 2, pp. 181–217, 2015.
11. S. Nahar, M. Azim, and C. Jubb, *The determinants of risk disclosure by banking institutions: evidence from Bangladesh*, Asian Review of Accounting, vol. 24, no. 4, pp. 426–444, 2016.
12. A. Al-Maghzom, K. Hussainey, and D. Aly, *The level of risk disclosure in listed banks: evidence from Saudi Arabia*, Corporate Ownership and Control, vol. 14, no. 1, pp. 175–194, 2016.
13. E. Sayed Abd Elghaffar, *Determining factors that affect risk disclosure level in Egyptian banks*, Banks and Bank Systems, vol. 14, no. 1, pp. 159–171, 2019.
14. E. Scannella, and S. Polizzi, *How to measure bank credit risk disclosure? Testing a new methodological approach based on the content analysis framework*, Journal of Banking Regulation, vol. 22, no. 1, pp. 73–95, 2021.
15. H. Qadiri, and S. Alsughayer, *Credit risk and disclosure behavior in the bank industry: evidence from Saudi Arabia*, International Journal of Accounting and Financial Reporting, vol. 11, no. 3, pp. 1–24, 2021.
16. M. M. A. Hussein, *The impact of disclosing bank credit risks on improving the financial performance of commercial banks in the Egyptian business environment: an applied study*, Journal of Business Research, vol. 8, no. 4, pp. 123–205, 2021.
17. Y. Altunbaş, A. Khan, and J. Thornton, *Disclosure and bank risk: evidence from European banks*, Finance Research Letters, vol. 58, 104499, 2023.
18. G. J. Lobo, C. Wang, and F. H. Wu, *IFRS 7 adoption and bank risk taking*, Journal of Accounting and Public Policy, vol. 46, 107225, 2024.
19. F. Ofria, and M. Mucciardi, *Government failures and non-performing loans in European countries: a spatial approach*, Journal of Economic Studies, vol. 49, no. 5, pp. 876–887, 2022.
20. A. Acheampong and T. Elshandidy, *Does soft information determine credit risk? Text-based evidence from European banks*, Journal of International Financial Markets, Institutions and Money, vol. 75, 101303, 2021.
21. M. M. Seliem, M. Amin, M. M. Abo El Nasr, E. A. Elnaggar, H. A. M. Khalifa, and M. A. A. Arab, *Performance of machine learning algorithms for credit risk prediction with feature selection*, Statistics, Optimization & Information Computing, vol. 14, pp. 311–328, 2025.
22. Elsegai, H., Al-Mutawaly, H., & Almongy, H. (2025). *Predicting the Trends of the Egyptian Stock Market Using Machine Learning and Deep Learning Methods*, Computational Journal of Mathematical and Statistical Sciences, 4(1), 186-221. doi: 10.21608/cjmss.2024.320645.1077
23. A. Faris and M. Elhachloufi, *Artificial intelligence and machine learning models for credit risk prediction in Morocco*, Statistics, Optimization & Information Computing, vol. 14, no. 4, pp. 1716–1740, 2025.

24. Z. Hjouji, I. Hasinat, and A. Hjouji, *A new method in machine learning adapted for credit risk prediction of bank loans*, Statistics, Optimization & Information Computing, vol. 13, no. 3, pp. 1209–1232, 2024.
25. E. M. Almetwally, I. Elbatal, M. Elgarhy, and A. R. Kamel, *Implications of machine learning techniques for prediction of motor health disorders in Saudi Arabia*, Alexandria Engineering Journal, vol. 127, pp. 1193–1208, 2025.
26. N. M. Al-Olaimat, H. S. Migdadi, A. Alrashidi, and S. Alzubaidi, *The Ujlayan–Dixit fractional Weibull distribution: statistical properties, estimation, and applications to engineering and biomedical data*, AIMS Mathematics, vol. 11, no. 10, pp. 35429–35468, 2026.
27. Abotaleb, M., Lim, W. H., Mishra, P., Qenawy, A., & ALmetwally, E. (2024). *Enhancing Stock Price Prediction Accuracy Using ARIMA and Advanced Greylag Goose Optimizer Algorithm*, Journal of Artificial Intelligence in Engineering Practice, 1(1), 49-65. doi: 10.21608/jaiep.2024.355004
28. N. M. Al-Olaimat, H. S. Migdadi, R. Saadeh, and M. Hafez, *On the two-parameter unit power Rayleigh distribution: properties, inference and applications*, Results in Control and Optimization, vol. 23, 100734, 2026.
29. Elbatal, I., Almetwally, E. M., Alaskar, M. Z., Alsaedi, B. S., Alzubaidi, S., & Yusuf, M. (2025). *Predictive analytics in accounting: Using regression models to forecast financial outcomes*, Alexandria Engineering Journal, 128, 484-492.
30. N. M. Al-Olaimat, H. S. Migdadi, H. A. Bayoud, and M. Z. Raqab, *Prediction methods for future record values from two-parameter Kies distribution*, Statistics, Optimization and Information Computing, vol. 15, no. 5, pp. 3381–3400, 2026.
31. S. Suthaharan, *Support vector machine*, in Machine Learning Models and Algorithms for Big Data Classification, Springer, Boston, MA, pp. 207–235, 2016.
32. D. Valkenborg, A. J. Rousseau, M. Geubbelmans, and T. Burzykowski, *Support vector machines*, American Journal of Orthodontics and Dentofacial Orthopedics, vol. 164, no. 5, pp. 754–757, 2023.
33. T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, and T. Zhou, *XGBoost: extreme gradient boosting*, R package version 0.4-2, vol. 1, no. 4, pp. 1–4, 2015.
34. M. Wiens, A. Verone-Boyle, N. Henscheid, J. T. Podichetty, and J. Burton, *A tutorial and use case example of the eXtreme gradient boosting (XGBoost) artificial intelligence algorithm for drug development applications*, Clinical and Translational Science, vol. 18, no. 3, e70172, 2025.
35. Y. C. Wu, and J. W. Feng, *Development and application of artificial neural network*, Wireless Personal Communications, vol. 102, no. 2, pp. 1645–1656, 2018.
36. S. Castillo-Girones, S. Munera, M. Martinez-Sober, J. Blasco, S. Cubero, and J. Gómez-Sanchis, *Artificial neural networks in agriculture, the core of artificial intelligence: what, when, and why*, Computers and Electronics in Agriculture, vol. 230, 109938, 2025.
37. Z. Zhang, *Introduction to machine learning: k-nearest neighbors*, Annals of Translational Medicine, vol. 4, no. 11, 218, 2016.
38. F. Acito, *k nearest neighbors*, in Predictive Analytics with KNIME: Analytics for Citizen Data Scientists, Springer Nature Switzerland, Cham, pp. 209–227, 2023.
39. G. Naidu, T. Zuva, and E. M. Sibanda, *A review of evaluation metrics in machine learning algorithms*, in Computer Science On-line Conference, Springer International Publishing, Cham, pp. 15–25, 2023.
40. M. Abonazel and A. Rabie, *The impact of using robust estimations in regression models: An application on the Egyptian economy*, Journal of Advanced Research in Applied Mathematics and Statistics, vol. 4, no. 2, pp. 8–16, 2019.
41. A. R. Kamel and M. R. Abonazel, *A simple introduction to regression modeling using R*, Computational Journal of Mathematical and Statistical Sciences, vol. 2, no. 1, pp. 52–79, 2023.
42. A. R. Kamel, M. R. Abonazel, and A. H. Youssef, *A novel robust M-estimator for the random-coefficients regression model: simulation and its application in energy management systems*, Computational Journal of Mathematical and Statistical Sciences, 2025.
43. M. R. Abonazel, A. R. Kamel, and A. H. Youssef, *New robust M-estimator for random-coefficients panel data model: algorithms, simulation, and application to insurance data*, Statistical Analysis and Data Mining: An ASA Data Science Journal, vol. 19, no. 3, e70079, 2026.
44. F. M. Alghamdi, A. R. Kamel, M. S. Mustafa, M. M. Bahloul, M. M. Alsolmi, and M. R. Abonazel, *A statistical study for the impact of REMS and nuclear energy on carbon dioxide emissions reductions in G20 countries*, Journal of Radiation Research and Applied Sciences, vol. 17, no. 3, 100993, 2024.
45. A. H. Youssef, M. R. Abonazel, and A. R. Kamel, *Efficiency comparisons of robust and non-robust estimators for seemingly unrelated regressions model*, WSEAS Transactions on Mathematics, vol. 21, no. 2, pp. 8–244, 2022.
46. R. Guerraoui, N. Gupta, and R. Pinot, *Robust Machine Learning*, Springer, Cham, Switzerland, 2024.