Developing an Early Identification Model for Bank Bankruptcy Using Unsupervised Learning

Ali Bagheri[1], Ali Zabihi [2,*] and Seyed Rasool Masoumi [3]

[1] Department of Accounting, Sar.C., Islamic Azad University, Sari, Iran;
[2] Department of Accounting, Sar.C., Islamic Azad University, Sari, Iran;
[3] Department of Accounting, Sar.C., Islamic Azad University, Sari, Iran;

* Correspondence: 5829778556@iau.ac.ir



Citation: Bagheri, A., Zabihi, A., & Masoumi, S. R. (2027). Developing an Early Identification Model for Bank Bankruptcy Using Unsupervised Learning. *Business, Marketing, and Finance Open*, 4(10), 1-20.

Received: 28 July 2026
Revised: 24 October 2026
Accepted: 01 November 2026
Initial Publication: 04 November 2026
Final Publication: 01 October 2027

Copyright: © 2027 by the authors. Published under the terms and conditions of Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License.

Abstract: The present study aimed to develop a model for the early identification of bank bankruptcy using unsupervised learning. To this end, a mixed qualitative–quantitative approach was employed. In the qualitative phase, indicators associated with the early identification of bank bankruptcy were extracted through in-depth semi-structured interviews with experts and subsequently validated based on their opinions. In the quantitative phase, data relating to 13 listed banks over the period 2016–2024 were collected and analyzed, yielding 117 bank-year observations. First, canonical correlation analysis was employed to examine the relationships between the extracted indicators and bank soundness assessment criteria. Subsequently, K-means clustering was used to classify the observations into five clusters. Local anomaly detection was then applied to identify abnormal observations and banks potentially exposed to bankruptcy risk. In addition, the model results were compared with the Altman Z-score, while the performance of the proposed approach was compared with the supervised learning algorithms of support vector machine and decision tree. The findings showed that, out of 117 observations, 92 observations, equivalent to 78.6%, were identified as anomalous. Of these, 42.7% were classified within the cluster of financially unhealthy banks or banks exposed to bankruptcy risk, whereas 57.3% were placed within the cluster of financially healthy banks. According to the Altman Z-score criterion, 13 observations, equivalent to 11.1%, were classified as financially healthy, whereas 104 observations, equivalent to 88.9%, were classified as financially distressed. Comparative results indicated that the proposed model achieved reported detection rates of 92.3% and 90.4% for healthy and unhealthy observations, respectively, and outperformed the support vector machine and decision tree algorithms in distinguishing financially healthy banks from banks exposed to bankruptcy risk. Overall, the findings indicate that integrating expert-based indicators with clustering and anomaly-detection techniques can provide a practical approach for the early identification of abnormal financial conditions in banks and can contribute to strengthening bank financial soundness monitoring systems.

Keywords: bank bankruptcy, early identification, unsupervised learning, anomaly detection, clustering.

1. Introduction

Banks occupy a central position in modern financial systems because they mobilize savings, allocate credit, facilitate payments, provide liquidity, and transmit monetary policy throughout the economy. The stability of banking institutions is therefore closely associated with the continuity of economic activity, investment, employment, and public confidence in the financial system. Unlike many nonfinancial firms, the failure of a bank can generate effects that extend well beyond its shareholders and creditors because of interconnected balance sheets, payment obligations, depositor behavior, and contagion across financial institutions. Consequently, bank distress may evolve from an institution-specific problem into a broader threat to financial stability, particularly

when several banks simultaneously experience deteriorating asset quality, liquidity shortages, capital inadequacy, or declining profitability. Historical experience has shown that financially distressed institutions may require extensive public intervention and emergency support, while such measures can impose considerable direct and indirect costs on governments and taxpayers [1]. For this reason, prudential supervision has progressively developed from a narrow emphasis on individual bank solvency toward a broader framework encompassing capital adequacy, risk management, liquidity, governance, and systemic resilience. The evolution of the Basel regulatory framework also reflects the increasing recognition that banking crises should, as far as possible, be prevented through continuous monitoring and timely identification of vulnerabilities rather than addressed only after financial distress has become severe [2]. Thus, identifying early warning signals of bank bankruptcy has become a critical issue for regulators, bank managers, investors, auditors, and other stakeholders concerned with preserving financial stability.

Bankruptcy in the banking sector is generally preceded by a progressive deterioration in several dimensions of financial health rather than by a single observable event. Weakening capital positions, rising non-performing loans, inadequate provisioning, excessive credit expansion, liquidity mismatches, declining earnings quality, increasing reliance on unstable funding sources, poor governance, and ineffective management may interact over time and gradually reduce a bank's ability to absorb shocks. Therefore, evaluating financial strength requires a multidimensional perspective. Comparative evidence concerning Islamic and conventional banks has demonstrated that overall bank financial strength is shaped by multiple interrelated financial and institutional characteristics rather than by a single accounting indicator [3]. Similarly, assessments of banking-sector efficiency indicate that differences in resource utilization, operational performance, profitability, and cost structures can significantly affect the resilience of banking institutions [4]. The identification of weak banks is consequently a multicriteria problem in which different dimensions of vulnerability must be considered simultaneously. Multicriteria decision-support approaches have, for example, been developed to identify weak banks on the basis of capital shortfalls and related measures of financial fragility, demonstrating the usefulness of integrating several dimensions of bank performance into early-warning systems [5]. Such findings suggest that an effective bankruptcy identification framework should capture the interaction among indicators of capital, assets, management, profitability, liquidity, market sensitivity, funding structure, and governance.

The importance of such monitoring is even greater in banking systems characterized by structural imbalances, concentrated exposures, regulatory constraints, macroeconomic instability, or persistent liquidity pressures. Under such circumstances, weaknesses in individual banks may reinforce vulnerabilities within the broader banking network. Research on systemic risk in Iran's banking network has emphasized the importance of institution-specific and systemic factors in explaining the propagation of financial risk across the banking sector [6]. In addition, Iranian research has attempted to identify specific characteristics of banks exposed to bankruptcy crises, indicating that early warning signals can emerge not only from conventional financial ratios but also from organizational, managerial, and governance-related factors [7]. The probability of financial distress may also be associated with broader dimensions of corporate behavior, reporting quality, and earnings characteristics. Evidence from firms listed on the Tehran Stock Exchange, for example, has examined the relationships among financial crisis, bankruptcy probability, social responsibility, and earnings quality, highlighting the potential importance of accounting information quality in assessing financial vulnerability [8]. In the banking context, these considerations imply that purely mechanical application of traditional financial ratios may fail to capture some of the qualitative and institutional warning signals that precede severe distress. An early-warning model may therefore benefit from

combining quantitative financial information with expert-informed indicators reflecting managerial changes, governance arrangements, funding behavior, and other warning signs that are not necessarily represented by conventional bankruptcy models.

Traditional bankruptcy prediction research has relied extensively on statistical classification methods and financial-ratio models. Such approaches have contributed substantially to the development of financial distress prediction; however, they also face several methodological limitations when applied to banks. First, the balance-sheet structure of banks differs fundamentally from that of industrial or commercial corporations. High leverage is intrinsic to banking activity, deposits constitute a major source of funding, liquidity transformation is central to the business model, and regulatory capital requirements play an essential role in determining institutional soundness. Consequently, bankruptcy criteria developed primarily for nonfinancial firms may not transfer directly to banking institutions. Second, the financial relationships preceding bank distress may be nonlinear and involve interactions that conventional linear techniques cannot easily capture. Third, failed-bank datasets are often highly imbalanced because the number of actual failures is generally small relative to the number of nonfailed institutions. Research has shown that class imbalance can materially affect bankruptcy prediction performance and may lead classifiers to favor the majority class while failing to detect the financially distressed cases that are of greatest practical importance [9]. These limitations have encouraged researchers to increasingly adopt machine-learning approaches that can detect complex patterns and improve classification performance under challenging data structures.

Machine-learning methods provide a flexible analytical framework for bankruptcy prediction because they can process high-dimensional data, capture nonlinear relationships, and uncover interactions among predictors without requiring the restrictive assumptions associated with some conventional statistical models. Studies applying machine learning to bank failure prediction have demonstrated that incorporating such methods into traditional financial-ratio analysis can improve predictive accuracy and enhance the identification of failing banks [10]. Support vector machines, decision trees, ensemble methods, neural networks, and other classification algorithms have consequently become prominent tools in financial distress research. Recent empirical studies have demonstrated the ability of advanced machine-learning techniques to distinguish failed from nonfailed banks using combinations of accounting and financial variables [11]. Similarly, research specifically focused on banking institutions has shown that advanced machine-learning approaches can enhance bankruptcy prediction and provide useful analytical support for risk assessment [12]. More generally, comparative studies of bankruptcy prediction algorithms indicate that machine-learning techniques can offer strong predictive performance when the relationships among financial indicators are complex and nonlinear [13]. Robust machine-learning models have also been proposed to improve stability and classification performance in bankruptcy prediction, particularly when traditional models are sensitive to data characteristics and model specification [14].

The development of deep learning has extended this line of research by enabling models to learn complex hierarchical representations from structured and unstructured financial information. Deep-learning models have been applied to bankruptcy prediction using textual disclosures, demonstrating that narrative corporate information can contain incremental signals of financial distress that may not be fully reflected in numerical financial ratios [15]. Subsequent research using the same general framework has further illustrated the predictive potential of textual disclosures and deep neural architectures for bankruptcy classification [16]. More recently, deep-learning algorithms have also been used to identify bankruptcy warning signs, reinforcing the argument that advanced computational techniques can contribute to earlier and more accurate identification of deteriorating financial conditions [17]. These developments have significantly expanded the analytical capacity of bankruptcy

prediction models. Nevertheless, most machine-learning studies remain based on supervised learning, which requires observations to be assigned predefined labels such as healthy, distressed, bankrupt, or nonbankrupt. The reliability of the resulting model therefore depends heavily on the quality and conceptual validity of these labels.

The dependence of supervised learning on predetermined outcome labels creates a particular challenge in banking research. Formal bankruptcy is relatively rare, regulatory intervention may occur before legal failure, and distressed banks may remain operational for extended periods because of recapitalization, liquidity support, mergers, regulatory forbearance, or public assistance. As a result, the distinction between a healthy bank and a bank that is economically distressed but not formally bankrupt may be ambiguous. Moreover, when historical failure labels are scarce, imbalanced, or affected by institutional interventions, supervised models may learn classification boundaries that do not adequately represent emerging or previously unseen forms of financial distress. Comprehensive evaluations of bankruptcy risk in the financial industry have therefore emphasized the importance of multidimensional risk assessment frameworks that combine financial indicators and advanced analytical methodologies rather than relying exclusively on a single bankruptcy criterion [18]. In this respect, unsupervised learning offers an important alternative because it does not require ex ante classification of observations. Instead, unsupervised algorithms identify latent structures, clusters, distances, density patterns, and anomalous observations directly from the characteristics of the data.

Unsupervised learning is particularly relevant to early bank-risk identification because early warning is fundamentally concerned with discovering unusual patterns before an adverse outcome has necessarily been observed. Clustering methods can group banks with similar financial profiles and thereby reveal distinct levels or patterns of financial health, while anomaly-detection techniques can identify observations that differ substantially from the local or global structure of the data. This capability is potentially valuable in banking supervision because an institution may exhibit an unusual combination of liquidity, profitability, funding, governance, or asset-quality indicators even when no formal bankruptcy label is available. Recent evidence has demonstrated the usefulness of unsupervised learning for monitoring bank risk across different countries, showing that data-driven methods can identify heterogeneous risk profiles and provide information relevant to continuous bank surveillance [19]. Such approaches shift the analytical focus from predicting a known historical label toward detecting deviations from normal financial behavior. This is especially important in rapidly changing financial environments, where the characteristics of future distressed banks may not precisely resemble those of institutions that failed in the past.

Cluster analysis is one of the principal unsupervised approaches that can be employed for this purpose. Algorithms such as K-means assign observations to groups by minimizing within-cluster distances and maximizing the differentiation among clusters. When a meaningful composite measure of bank financial soundness is available, clustering can be used to organize bank-year observations into distinct groups representing different degrees or patterns of financial health. Density-based clustering methods provide a complementary perspective because they identify clusters according to the concentration of observations in multidimensional space and can distinguish observations that do not belong strongly to dense groups. In banking applications, this feature is useful because financially abnormal institutions may occupy sparse areas of the data distribution. Anomaly-detection methods provide an additional layer of analysis by quantifying the extent to which individual observations diverge from neighboring observations. A local approach is particularly attractive because an observation can be considered anomalous relative to banks with broadly comparable characteristics even if it is not an extreme observation in the entire dataset. Accordingly, the combination of clustering and local anomaly detection can provide a flexible

framework for identifying unusual financial conditions without presupposing a fixed boundary between healthy and bankrupt banks.

The predictive usefulness of machine learning in banking is not restricted to bankruptcy itself. Hybrid machine-learning approaches have also been proposed for predicting banking liquidity risk, reflecting the growing use of computational models to identify financial vulnerabilities before they develop into severe crises [20]. Liquidity risk is particularly relevant because banks perform maturity transformation by funding longer-term assets with shorter-term liabilities, making them inherently vulnerable to mismatches in cash-flow timing and changes in depositor behavior. Excessive loan-to-deposit ratios, negative maturity gaps, dependence on costly or external financing, and aggressive deposit-mobilization strategies may therefore represent important early warning signals. When these liquidity-related indicators are considered alongside profitability, provisioning, asset quality, capital adequacy, and governance indicators, they can provide a richer representation of the processes through which bank soundness deteriorates. The development of an effective early-identification system consequently requires careful selection of indicators rather than indiscriminate inclusion of a large number of financial variables.

Expert knowledge can play an important role in this selection process. Statistical relationships identified from historical data are valuable, but they may not fully represent institutional practices, regulatory concerns, or behavioral warning signs observed by banking professionals. Qualitative investigation can identify indicators that experienced practitioners regard as significant before severe financial deterioration becomes visible in conventional accounts. Previous Iranian research based on thematic analysis has demonstrated the feasibility of systematically identifying indicators of banks exposed to bankruptcy crisis through expert knowledge [7]. Combining such qualitatively identified indicators with quantitative machine-learning techniques can therefore improve the conceptual grounding of the resulting model. Rather than treating the predictor set merely as a collection of available accounting ratios, an expert-informed framework can incorporate variables related to management turnover, auditor changes, governance structure, deposit policies, funding dependence, and other signals of organizational instability. This mixed approach may be particularly appropriate in banking systems where institutional characteristics and policy interventions exert a strong influence on observed financial outcomes.

Another important requirement is the validation of any proposed unsupervised classification against established measures of financial health and alternative predictive approaches. Because unsupervised learning generates clusters or anomaly scores without known target labels, the practical interpretation of the resulting groups must be supported by external criteria. The CAMELS framework provides one relevant reference structure because it evaluates bank condition through capital adequacy, asset quality, management, earnings, liquidity, and sensitivity to market risk. These dimensions reflect the multidimensional nature of bank soundness and can be employed to examine whether newly developed indicators have meaningful relationships with recognized prudential criteria. Canonical correlation analysis is especially useful in this context because it evaluates the association between two sets of variables and identifies linear combinations that maximize the correlation between them. By linking expert-derived indicators to CAMELS dimensions, a composite financial soundness criterion can be established before clustering is performed. Such an approach is consistent with the broader literature emphasizing comprehensive rather than single-dimensional assessment of bank strength [3, 5].

External comparison with bankruptcy scoring models can provide an additional validation mechanism, although such comparisons must be interpreted carefully in banking applications. The Altman Z-score is among the most widely recognized measures of corporate financial distress, but its original structure was not designed specifically for commercial banks. Nevertheless, it can serve as a benchmark for evaluating whether observations

identified as financially abnormal by an unsupervised procedure also exhibit characteristics associated with financial distress under a conventional scoring approach. More importantly, comparing the performance of an unsupervised method with supervised classifiers such as support vector machines and decision trees can clarify whether predefined labels are necessary for achieving satisfactory discrimination. Existing evidence has shown that machine-learning classifiers can perform effectively in bankruptcy prediction [10, 11, 13], but the question remains whether an unsupervised framework can provide comparable or superior detection when bank distress labels are incomplete or uncertain.

This issue is particularly significant from a supervisory and managerial perspective. Early-warning systems are most valuable when they identify deteriorating conditions sufficiently early to permit corrective intervention. If warning signals are detected only after formal insolvency or severe regulatory deterioration, the practical value of the model is substantially reduced. Banking authorities may use early identification to intensify supervisory review, require corrective capital or liquidity plans, restrict risky expansion, evaluate governance deficiencies, or initiate resolution planning. Bank managers can similarly use such systems to detect changes in financial structure before they become critical. Investors, creditors, and auditors may benefit from additional evidence concerning unusual risk profiles. The costs associated with delayed recognition can be substantial, as demonstrated by the need for large-scale intervention during banking crises [1]. Consequently, an effective early-identification system should not merely maximize retrospective classification accuracy; it should also provide interpretable information concerning the financial and governance indicators that distinguish normal from abnormal bank conditions.

Despite the growing literature on bankruptcy prediction, several research gaps remain. First, much of the existing evidence focuses on nonfinancial corporations, whereas banks possess distinct financial structures and regulatory constraints. Second, many banking studies employ predefined indicators without systematically integrating expert knowledge into the model-development process. Third, supervised learning continues to dominate the field despite the scarcity and ambiguity of actual bank-failure labels. Fourth, studies frequently evaluate a single machine-learning method without integrating indicator validation, clustering, anomaly detection, and benchmarking within a unified framework. Fifth, the specific institutional characteristics of banking systems such as Iran's may limit the direct transferability of models developed in other countries. The presence of systemic vulnerabilities in Iran's banking network further reinforces the need for context-sensitive monitoring approaches [6]. Recent research on bankruptcy warning signs and machine learning demonstrates substantial methodological progress [17, 19], yet there remains a need for models capable of identifying abnormal bank conditions without relying exclusively on previously observed bankruptcy cases.

Accordingly, a framework that begins with qualitative identification of warning indicators, validates those indicators against established dimensions of bank soundness, constructs a composite financial health measure, uses unsupervised clustering to identify latent groups, applies local anomaly detection to distinguish unusual observations, and then compares its classifications with conventional and supervised approaches can address several limitations of previous research simultaneously. Such a framework recognizes that bank bankruptcy is not simply a binary event but the potential endpoint of a gradual and multidimensional deterioration process. It also acknowledges that anomalous behavior may occur within banks that appear broadly healthy according to conventional criteria and that early-warning systems should be capable of detecting such localized abnormalities. By combining expert judgment with data-driven analytical techniques, the proposed approach can strengthen the interpretability and practical relevance of machine-learning applications in bank supervision while reducing dependence on scarce bankruptcy labels. It may therefore provide a more flexible basis for monitoring financial

health, recognizing emerging vulnerabilities, and supporting preventive decision-making by regulators and bank managers.

Therefore, the aim of the present study is to develop and evaluate an early identification model for bank bankruptcy by integrating expert-derived indicators, canonical correlation analysis, unsupervised clustering, and local anomaly detection and by comparing its performance with the Altman Z-score and supervised machine-learning algorithms.

2. Methods and Materials

The present study is applied–developmental in terms of purpose and adopts a mixed-methods design comprising qualitative and quantitative phases. In the qualitative phase, preliminary indicators of bank bankruptcy are extracted through a review of documents and scientific literature and through semi-structured interviews with experts in finance, accounting, and banking. The indicators are subsequently evaluated and finalized using qualitative analysis and the Delphi technique. Experts are selected through purposive and snowball sampling on the basis of their relevant expertise and professional experience. In the quantitative phase, the financial data of 13 banks listed on the Tehran Stock Exchange during 2016–2024, comprising 117 bank-year observations, are examined. First, canonical correlation analysis is employed to investigate the relationships between the indicators and the dimensions of the CAMELS framework. Subsequently, the K-means clustering algorithm and a local anomaly-detection method are applied to group the observations and identify abnormal cases. Finally, the results of the proposed model are compared with the Altman Z-score and the support vector machine and decision tree algorithms. Given the limitations associated with the direct application of the Altman criterion to banks, its results are interpreted with caution.

3. Findings and Results

In the qualitative phase of the study, semi-structured interviews were conducted with 15 experts and specialists in banking, university faculty members, and financial researchers to identify and finalize the preliminary indicators of bank bankruptcy. Participants were selected using purposive and snowball sampling, and the interview process continued until theoretical saturation was achieved. The collected viewpoints formed the basis for extracting, examining, and categorizing the preliminary indicators of bank bankruptcy, and 18 preliminary indicators were identified through thematic analysis of the interviews conducted with the experts. Following completion of this process and extraction of the preliminary indicators from the expert interviews, the Delphi technique was employed to ensure the internal validity of the data and to ascertain that the researchers' interpretations of participants' statements were consistent with their intended meanings. Following three Delphi rounds, 15 indicators were confirmed as the final indicators of the study. Subsequently, through continuous comparison of the indicators and examination of their similarities and conceptual relationships, related indicators were classified into higher-order organizing themes. The results of this classification and the expert agreement ratios for the corresponding organizing themes are presented in Table 1.

Table 1. Formation of Organizing Themes Based on the Preliminary Indicators

Organizing Theme	Preliminary Indicator	Agreement Ratio
Financial factors	Loan-to-deposit ratio (LTD)	0.80
	Capital adequacy ratio (CAR)	0.93
	Non-performing loan ratio (NPL)	0.93

	Loan-loss reserve coverage ratio	1.00
	Return on assets (ROA)	1.00
	Balance sheet growth	1.00
	Dependence on external financing	0.80
	Cost-to-income ratio	0.87
	Non-operating income ratio	0.87
	Negative maturity gap	0.80
Governance and policy factors	Change of chief executive officer in two consecutive years	1.00
	Change of independent auditor	0.80
	Proportion of non-executive board members	1.00
	Chief executive officer's managerial experience	—
	Introduction of new deposit-attraction schemes offering higher interest rates	0.93

The banks' financial soundness index was constructed using the indicators identified in the qualitative phase of the study. For this purpose, canonical vectors and canonical correlation coefficients between the identified indicators and the six CAMELS criteria were estimated. Subsequently, K-means cluster analysis was used to classify the banks into five clusters, consistent with the CAMELS classification system. Given the six CAMELS criteria and the 10 indicators proposed in the study, six canonical vectors and coefficients—the minimum number of indicators across the two groups—were estimated. Bartlett's test was then employed to determine the statistical significance of these canonical vectors and canonical correlation coefficients. The results are presented in Table 2.

Table 2. Canonical Correlation Coefficients and Bartlett's Significance Test

Canonical Vector	Canonical Correlation Coefficient	R^2	Bartlett's Chi-Square	Significance
1	0.776444	0.60287	191.54	0.0001
2	0.689626	0.47558	92.265	0.0001
3	0.312527	0.09767	22.877	0.8821
4	0.248853	0.061928	11.828	0.9441
5	0.188758	0.035630	4.9556	0.9594
6	0.098847	0.0097708	1.0555	0.9580

The results presented in Table 2 indicate that, among the six estimable canonical vectors and canonical correlation coefficients between the CAMELS indicators and the 10 study indicators, two canonical correlation vectors had significance levels below .05 and were therefore statistically capable of explaining the linear relationships between the indicators of the two sets. Of these two canonical functions, the first canonical vector, with a canonical correlation coefficient of 0.7764, demonstrated the greatest ability to establish an association between the two groups of indicators. Based on these findings, the canonical coefficients for the first fitted canonical vector corresponding to the 10 study indicators and CAMELS indicators are presented in Table 3. These coefficients represent the optimal linear combination of the CAMELS indicators and the 10 study indicators that produces the strongest association between the CAMELS criterion and the financial soundness criterion developed in this study.

Table 3. Canonical Vector Coefficients for the Study and CAMELS Indicators

Indicator	Canonical Vector Coefficient X	Canonical Vector Coefficient Y
Loan-to-deposit ratio	5.1936	—
Reserve coverage ratio	-2.2590	—
Balance sheet growth	-0.3065	—
Negative maturity gap	1.8531	—
Non-operating income ratio	-2.5555	—

Change of chief executive officer	0.5792	—
New deposit-attraction scheme	0.5309	—
Change of auditor	0.2261	—
Proportion of non-executive board members	-4.1043	—
Chief executive officer's managerial experience	-0.0239	—
Capital adequacy	—	-1.9723
Asset quality	—	-14.0475
Management	—	-0.8526
Earnings	—	-14.1383
Liquidity	—	-10.5677
Sensitivity	—	-1.3879

Accordingly, the linear combinations **U** and **V**, derived from the two groups of X and Y indicators using the coefficients reported in Table (4-9), generate vectors with a correlation coefficient of 0.7764 between the two groups of indicators. The vector corresponding to the X indicators is introduced in the present study as the financial soundness index.

Based on the financial soundness criterion **U**, calculated from the canonical vector comprising the 10 study indicators, banks were clustered using the K-means algorithm into five clusters, consistent with the standard number of clusters in the CAMELS framework. The cluster centroids obtained from this analysis are presented in Table 4.

Table 4. Cluster Centroids Based on the U Financial Soundness Criterion

Cluster	Centroid
Cluster 1	-1.3693506
Cluster 2	1.3128178
Cluster 3	-0.4732817
Cluster 4	0.3685797
Cluster 5	2.4293405

Accordingly, to determine the cluster membership of each bank, the Euclidean distance between the bank's **U** vector value and the cluster centroids can be calculated. This procedure determines the cluster to which each bank belongs and provides the basis for clustering the banks according to the **U** financial soundness index. Figure 1 presents the silhouette plot corresponding to the bank clusters formed on the basis of the **U** index. This index is determined on the basis of the cohesion of observations within each cluster and their separation or distinguishability from observations belonging to other clusters. Silhouette values range from -1 to +1, with larger positive values indicating better separation of observations across clusters.

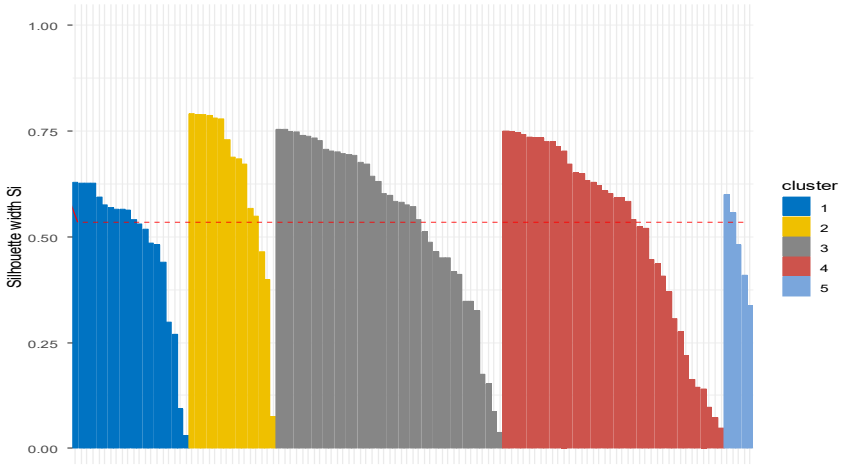


Figure 1. Silhouette Index for Bank Clustering Based on the U Financial Soundness Index

In determining the clusters using the DBSCAN method, tuning the algorithm's hyperparameters is the first priority. For this purpose, the k-nearest-neighbor distance plot of the observations was used. This plot is employed to determine the epsilon parameter in the DBSCAN algorithm.

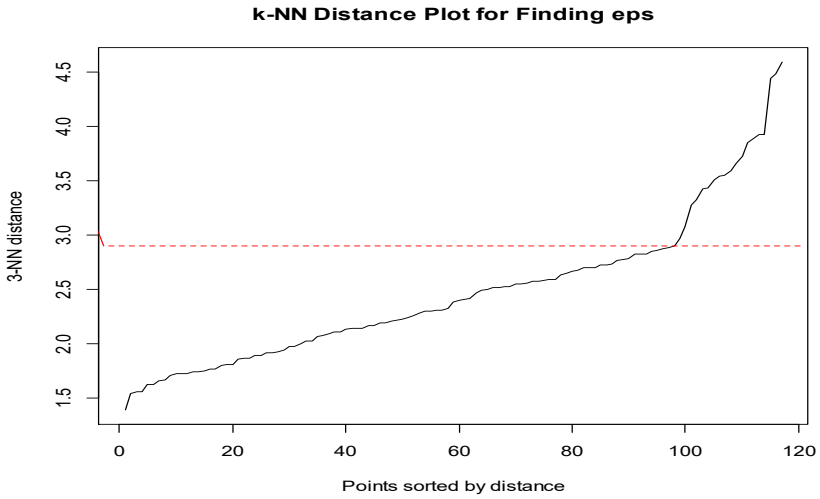


Figure 2. K-Nearest-Neighbor Distance Plot of the Observations

In this plot, the elbow point—that is, the point at which the curve begins to exhibit exponential growth and a marked change in its upward slope—determines the optimal value of the model hyperparameter. As shown in Figure 1, the optimal hyperparameter value in this analysis was estimated to be 2.96. After substituting this hyperparameter value into the DBSCAN algorithm, two clusters were identified for the banks. Among the observations, only 2.5% of the total observations were assigned to the second cluster, resulting in a substantial decline in predictive accuracy when these cluster labels were applied to the data. Figure 3 illustrates the clustering of these observations in this analysis.

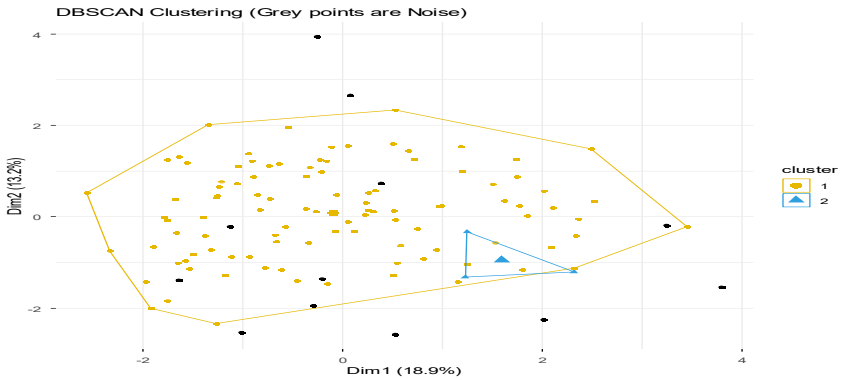


Figure 3. Bank Clustering Using the DBSCAN Method

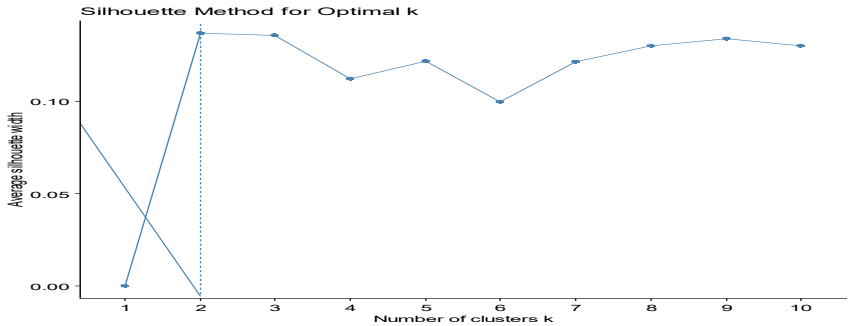


Figure 4. Cluster Quality and Density Based on the Average Silhouette Width

As shown in Figure 4, the average silhouette width reached its highest possible value for a two-cluster solution and declined as the number of clusters increased. This finding indicates that, according to this criterion, two clusters constitute the optimal number of clusters for the analysis of the study data. To identify anomalous banks, the local reachability density of each observation within each cluster and the Local Outlier Factor (LOF) based on this measure were employed. Based on the calculated LOF values and an anomaly-detection threshold equal to 1, the total number of anomalous observations across both clusters was identified as presented in Table 5.

Table 5. Frequency Distribution of Anomalous Observations Across Clusters

Cluster	Frequency	Percentage
Cluster 1	67	57.3%
Cluster 2	25	42.7%
Total anomalous observations	92	78.6%

According to the results presented in this table, 92 of the 117 observations were identified as anomalous, representing 78.6% of all observations examined. Of these observations, 42.7% were located in the cluster of financially unhealthy banks or banks exposed to bankruptcy risk, whereas 57.3% were located in the cluster of financially healthy banks. Figure 5 presents the box plot of local anomaly values for each cluster.

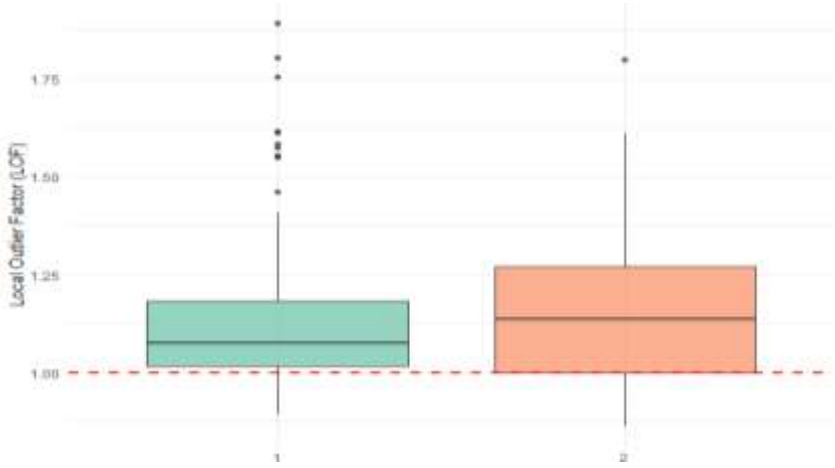


Figure 5. Box Plot of Local Anomaly Values of Observations by Cluster

Given the premise that the anomalous observations identified through this analysis represent banks exposed to bankruptcy risk based on the 10 study indicators, the performance of this algorithm in detecting bank bankruptcy was validated. For this purpose, the bankruptcy status of the banks was first determined using the Altman Z-score criterion, after which the ability of the study's unsupervised algorithm to discriminate among banks was evaluated against the labels derived from the Altman Z-score. Table 6 presents the confusion matrix for predictions generated by this algorithm.

Table 6. Confusion Matrix for Predicting Bank Status Using the Unsupervised Algorithm

Bank Status Based on Algorithmic Detection	Healthy According to Altman Criterion	Unhealthy According to Altman Criterion
Healthy	12	10
Unhealthy	1	94

Based on these results, under the Altman Z-score criterion, 13 bank-year observations were classified as financially healthy, representing 11.1% of the observations, whereas 104 observations were classified as being in a state of financial distress, representing 88.9%. Among the healthy banks, the unsupervised algorithm employed in this study correctly predicted 92.3%, equivalent to 12 observations, as healthy. Among the unhealthy banks, the algorithm correctly identified 90.4% as being in an unhealthy condition. These findings indicate the satisfactory performance of the proposed method in identifying and distinguishing financially healthy banks from banks exposed to risk using the study's financial indicators. The area under the ROC curve in this analysis was 0.913, further indicating the favorable performance of this method in identifying banks' financial soundness status. In detecting the financial health or financial distress status of banks using this algorithm, the maximum decision-tree depth was set at six levels, while the other model hyperparameters were identified and estimated through implementation packages available in R software. In this analysis, four decision-tree levels were generated to determine banks' financial soundness or financial distress status. Across these levels, only four of the 10 indicators examined served as the basis for decision-making. Figure 6 illustrates the decision-tree structure used to identify banks' financial soundness or financial distress status under this method.

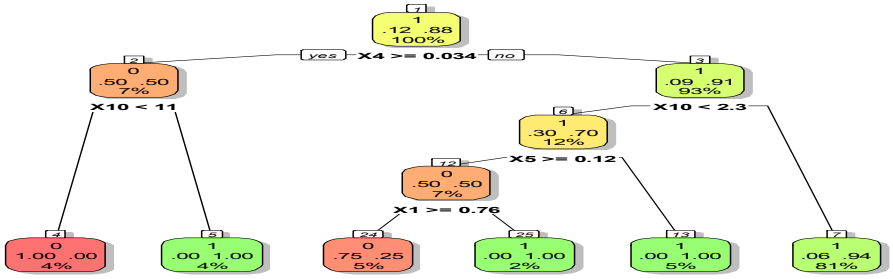


Figure 6. Decision Tree for Interpreting Bank Clusters

Based on the results shown in this figure, the loan-to-deposit ratio, negative maturity gap, non-operating income ratio, and chief executive officer's managerial experience were identified as the decision variables used to determine banks' status. The selection of variables at the decision-tree levels was based on their estimated importance according to the Gini coefficient. The corresponding values for these indicators are presented in Table 7.

Table 7. Gini Importance Coefficients of Indicators Used in Constructing the Decision Tree

Indicator	Importance Coefficient
Negative maturity gap (X4)	4.16389
Non-operating income ratio (X5)	4.150977
Chief executive officer's managerial experience (X10)	4.004885
Loan-to-deposit ratio (X1)	3.8

According to the results presented in Table 7 and the importance coefficients of the variables selected in the decision-tree model, the negative maturity gap had the highest importance among the selected variables and was therefore included as the decision variable at the first level of the decision tree. The loan-to-deposit ratio had the lowest importance among these variables and was accordingly placed at the final level of the decision tree. The results of bank-status classification based on the confusion matrix for this prediction method are presented in Table 8.

Table 8. Confusion Matrix for Predicting Bank Status Using the Decision-Tree Algorithm

Bank Status Based on Algorithmic Detection	Healthy According to Altman Criterion	Unhealthy According to Altman Criterion
Healthy	8	13
Unhealthy	5	91

Based on the results presented in Table 8, among banks classified as healthy according to the Altman Z-score criterion, the decision-tree algorithm correctly predicted 61.5%, equivalent to 8 observations, as healthy. Among the unhealthy banks, the algorithm correctly identified 87.5% of those in an unhealthy condition. These findings indicate the satisfactory performance of this method in identifying and distinguishing unhealthy banks using the financial indicators examined in the study. The area under the ROC curve in this analysis was estimated at 0.769, indicating that the classifications obtained using this approach were superior to those based on the support vector machine, but remained weaker than those obtained using the unsupervised learning algorithm.

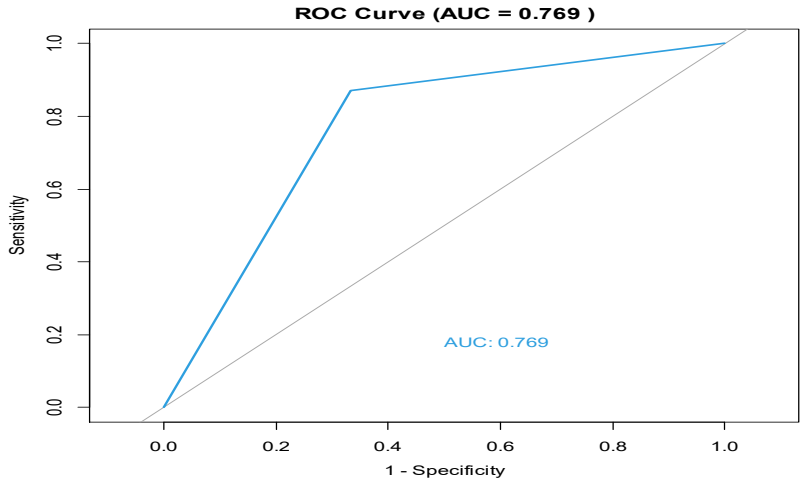


Figure 7. ROC Curve for Detecting Banks' Financial Soundness Status Using the Decision-Tree Algorithm

Based on these results, regarding the third research question, it can be concluded that the unsupervised algorithm based on the local anomaly characteristics of the observations performed substantially better in distinguishing financially healthy banks from banks exposed to bankruptcy risk than supervised learning algorithms, namely the support vector machine and decision tree. The high predictive accuracy obtained on the basis of the 10 indicators identified in the study further demonstrates that these indicators possess considerable discriminatory power for distinguishing financially healthy banks from banks exposed to bankruptcy risk.

4. Discussion and Conclusion

The present study developed and evaluated an early identification model for bank bankruptcy by integrating expert-derived indicators with canonical correlation analysis, K-means clustering, DBSCAN, local anomaly detection, and comparative supervised machine-learning algorithms. The findings showed that the proposed set of indicators was meaningfully associated with established dimensions of bank financial soundness and that the unsupervised framework demonstrated a strong capacity to distinguish healthy banks from banks exposed to financial distress. In the qualitative phase, 18 preliminary indicators were initially extracted from expert interviews, of which 15 were retained after three rounds of the Delphi process. These indicators were subsequently organized into financial and governance/policy-related dimensions. The inclusion of both financial and governance variables is important because bank distress rarely emerges from a single accounting weakness; rather, it is usually the product of interacting deficiencies in capital structure, liquidity, asset quality, profitability, funding dependence, governance, and management. This interpretation is consistent with prior research emphasizing that overall bank financial strength is inherently multidimensional and cannot be adequately represented by one isolated ratio [3]. It is also in line with findings indicating that banks exposed to bankruptcy crises can be identified through a combination of financial, managerial, and institutional indicators rather than through conventional financial variables alone [7]. The present results therefore support the argument that early-warning systems should combine quantitative financial measures with governance-related signals such as chief executive officer turnover, auditor change, board structure, management experience, and aggressive deposit-attraction policies.

The canonical correlation analysis provided further evidence supporting the construct validity of the selected indicators. Among the six possible canonical functions estimated between the 10 study indicators and the six CAMELS dimensions, the first two canonical correlations were statistically significant, with the first function

producing a canonical correlation of 0.7764 and explaining approximately 60.3% of the shared variance. This relatively strong association indicates that the proposed indicators capture a considerable proportion of the financial-health information represented by the CAMELS framework. In practical terms, the finding suggests that variables such as the loan-to-deposit ratio, reserve coverage, balance-sheet growth, negative maturity gap, non-operating income, managerial changes, auditor changes, board composition, and managerial experience are meaningfully connected to the established banking dimensions of capital adequacy, asset quality, management, earnings, liquidity, and sensitivity. This result aligns with prior research emphasizing the usefulness of multidimensional bank assessment frameworks in evaluating institutional strength [3]. It also supports the multicriteria perspective advanced in earlier studies, according to which weak banks are more accurately identified when several dimensions of vulnerability are considered simultaneously rather than assessed independently [5]. The findings are additionally compatible with evidence from banking-efficiency research suggesting that differences in cost structures, profitability, and operational resource allocation are closely related to the resilience and performance of banking systems [4]. Therefore, the canonical correlation results indicate that the selected indicators are not arbitrary predictors but are substantively linked to conventional prudential measures of bank soundness.

The K-means clustering results also demonstrated that the financial soundness index derived from the first canonical function could be used to differentiate bank-year observations into distinct groups. The five-cluster solution was initially selected to correspond to the conventional CAMELS classification structure, thereby providing a structured representation of varying levels of financial health. However, subsequent DBSCAN and silhouette analyses suggested that a two-cluster structure provided the strongest separation in the observed data. The higher average silhouette width for two clusters indicates that the data may be more naturally divided into two broad financial regimes, namely relatively healthy banks and banks exhibiting higher levels of vulnerability or distress. This finding is important because it suggests that although supervisory frameworks may conceptually distinguish multiple grades of financial condition, the empirical structure of the sample may be dominated by a more fundamental healthy-versus-risky distinction. Similar concerns have been raised in machine-learning studies, where the effective classification boundary is often shaped by the actual distribution of the data rather than by externally imposed categories [13]. The result is also consistent with recent unsupervised-learning research showing that latent bank-risk groupings can emerge directly from financial data without requiring predefined failure labels [19]. In this respect, the present study adds evidence that unsupervised clustering can reveal the intrinsic structure of banking risk and may provide a useful complement to conventional regulatory classification systems.

A particularly important result concerned the application of local anomaly detection. Based on the Local Outlier Factor threshold, 92 of the 117 observations, equivalent to 78.6%, were identified as anomalous. Among these, 42.7% belonged to the cluster interpreted as unhealthy or exposed to bankruptcy risk, whereas 57.3% belonged to the cluster interpreted as healthy. At first glance, the identification of anomalies within the healthy cluster may appear counterintuitive; however, from an early-warning perspective, this result is analytically meaningful. A local anomaly does not necessarily indicate current bankruptcy but rather signifies that an observation differs from neighboring observations with otherwise similar profiles. Consequently, a bank may still be broadly classified as healthy while exhibiting unusual combinations of funding, liquidity, profitability, governance, or balance-sheet characteristics that warrant closer monitoring. This is precisely where local anomaly detection can offer advantages over simple binary classification. Unsupervised methods are capable of detecting previously unrecognized or

emerging patterns of financial vulnerability without requiring that such patterns already correspond to historical failure cases [19]. This interpretation is also consistent with comprehensive bankruptcy-risk frameworks that emphasize the need to evaluate combinations of financial-risk signals rather than relying on a single threshold or deterministic classification rule [18]. Thus, the relatively high proportion of anomalous observations may reflect substantial heterogeneity within the sample and the sensitivity of the local method to deviations that are not necessarily captured by conventional distress labels.

The validation results based on the Altman Z-score further supported the discriminatory capacity of the proposed approach. According to the Altman criterion, 13 observations, equivalent to 11.1%, were classified as financially healthy, whereas 104 observations, equivalent to 88.9%, were classified as financially distressed. Against these benchmark labels, the unsupervised algorithm correctly identified 12 of the 13 healthy observations, corresponding to a detection rate of 92.3%, and correctly classified 94 of the 104 unhealthy observations, corresponding to 90.4%. The area under the ROC curve was 0.913, indicating a high level of discrimination. These results suggest that the proposed model not only detects structural anomalies but also produces classifications that are strongly consistent with an established financial-distress benchmark. The strong performance of the model is compatible with the broader bankruptcy-prediction literature showing that machine-learning approaches can substantially improve the identification of financially distressed entities compared with traditional linear models [10]. Similar studies have reported that machine-learning models are capable of capturing nonlinear interactions among financial ratios and improving prediction accuracy [12, 13]. The present findings extend this evidence by showing that unsupervised learning can also generate high discriminatory performance, even without being trained directly on predefined bankruptcy labels.

The strong performance of the unsupervised model is especially relevant given the structural limitations of supervised learning in the banking sector. Supervised techniques require reliable labels indicating which banks are healthy and which are bankrupt or distressed. However, formal bank failure is relatively uncommon, and distressed banks may remain operational for long periods due to regulatory intervention, liquidity assistance, recapitalization, mergers, or other support measures. Therefore, the observable label of "bankrupt" may not fully represent the underlying continuum of bank fragility. The banking literature has demonstrated that public intervention can materially alter the trajectory of distressed financial institutions, making formal failure an imperfect representation of underlying financial weakness [1]. Moreover, the evolution of international banking regulation has increasingly emphasized early supervisory intervention rather than waiting for formal insolvency [2]. In this context, the ability of an unsupervised model to detect abnormal financial profiles without relying on historical bankruptcy labels is particularly valuable. The findings therefore support recent work demonstrating that unsupervised learning can be effectively used to monitor bank risk across heterogeneous financial systems [19].

The comparison with the decision-tree algorithm further highlighted the superiority of the proposed method. The decision tree correctly identified only 61.5% of the healthy observations and 87.5% of the unhealthy observations, with an area under the ROC curve of 0.769. Although this represents an acceptable level of classification performance, it was clearly inferior to the unsupervised model. The findings therefore suggest that the data structure may contain local patterns and heterogeneous relationships that are not fully represented by the hierarchical partitioning rules generated by a decision tree. Prior studies have shown that machine-learning methods can differ substantially in their predictive performance depending on the characteristics of the bankruptcy dataset, the predictor structure, and the degree of class imbalance [9]. Decision-tree methods are interpretable and operationally useful, but they may be sensitive to sample size, variable selection, and split criteria. More robust

machine-learning frameworks have therefore been proposed to improve classification stability and generalization [14]. The superior performance of the unsupervised model in the present study supports the view that no single supervised classifier should be assumed to be optimal and that data-driven anomaly detection may provide important advantages in banking applications.

The decision-tree findings nevertheless offered valuable interpretive insight by identifying four key variables as the main decision factors: negative maturity gap, non-operating income ratio, chief executive officer's managerial experience, and loan-to-deposit ratio. Among these, the negative maturity gap had the highest Gini importance, suggesting that liquidity structure and maturity mismatch are particularly influential in distinguishing healthy from financially vulnerable banks. This finding is theoretically plausible because banks rely on maturity transformation and are therefore highly exposed to mismatches between the timing of asset inflows and liability obligations. A persistent negative maturity gap may indicate that short-term obligations exceed available short-term funding capacity, increasing the probability of liquidity stress. The importance of liquidity risk in bank distress is also consistent with studies developing machine-learning models specifically for banking liquidity-risk prediction [20]. The loan-to-deposit ratio likewise reflects the relationship between lending activity and relatively stable deposit funding, and excessive values may signal increased dependence on alternative or potentially costly funding sources. These findings reinforce the importance of liquidity-related variables in early-warning systems and suggest that they should be monitored jointly with profitability and governance indicators.

The importance of the non-operating income ratio also deserves attention. A high dependence on non-operating income may indicate that a bank's core intermediation activities are not generating sufficient sustainable earnings, potentially leading management to rely on irregular or nonrecurring income sources. Such a pattern may conceal deterioration in the underlying quality of banking operations. Previous research has shown that earnings quality and financial distress are closely related and that the interpretation of financial health can be improved by considering the quality and sustainability of reported income rather than relying exclusively on aggregate profitability [8]. In the present study, the role of non-operating income suggests that profitability should be evaluated not only quantitatively but also in terms of its composition and persistence. This is compatible with broader evidence that the financial strength of banks depends on multiple interacting dimensions, including earnings, asset quality, and management efficiency [3].

The identification of chief executive officer managerial experience as an important variable further underscores the relevance of governance and managerial capability in bank financial health. Managerial experience may affect the quality of lending decisions, liquidity management, risk assessment, strategic responses to macroeconomic shocks, and compliance with regulatory requirements. Poor managerial decisions can intensify asset-quality problems, encourage excessive credit expansion, or delay corrective action in periods of financial deterioration. The presence of management-related variables among the most important predictors therefore supports the inclusion of governance and organizational factors in bank distress models. This finding is consistent with prior thematic research identifying managerial and governance indicators among the warning signs of banks exposed to bankruptcy crises [7]. It also aligns with multidimensional banking frameworks such as CAMELS, in which management quality constitutes an independent and important component of supervisory assessment.

The overall superiority of the proposed unsupervised framework over the supervised alternatives is consistent with broader developments in bankruptcy prediction research. Studies using support vector machines, decision trees, neural networks, and other machine-learning techniques have shown that nonlinear computational methods can outperform traditional statistical approaches in complex financial datasets [10, 11]. More recent research has

extended this logic to deep learning, demonstrating that models capable of extracting complex latent representations can improve bankruptcy-warning performance [17]. Deep-learning studies using textual disclosures have similarly shown that relevant distress signals may be embedded in information that conventional ratio-based models do not fully capture [15, 16]. Although the present study did not use deep learning or textual data, its findings support the same underlying principle: financial distress is multidimensional, nonlinear, and not fully observable through predetermined labels. The use of unsupervised learning is therefore appropriate when the objective is to detect unusual financial structures rather than merely reproduce historical bankruptcy classifications.

The results also have implications for the Iranian banking system. Previous research has demonstrated the presence of systemic-risk drivers within Iran's banking network and has highlighted the possibility that vulnerabilities in individual institutions can propagate through the broader financial system [6]. Under such circumstances, early identification of abnormal financial patterns becomes especially important. A model that detects anomalies in liquidity, profitability, funding dependence, governance, or other key indicators can support earlier regulatory attention before institution-specific weaknesses become systemic. The findings of the present study therefore complement research on systemic banking risk by providing an institution-level detection mechanism that may be used as part of a wider supervisory framework. At the same time, the strong discrimination achieved by the model supports the argument that advanced computational tools can be effectively adapted to the specific characteristics of domestic banking environments rather than imported mechanically from studies conducted in other financial systems.

Overall, the findings suggest that combining expert knowledge with unsupervised machine learning can provide a robust and interpretable approach to the early identification of bank bankruptcy risk. The expert-derived indicators ensured that the model incorporated contextually meaningful financial and governance warning signs, while canonical correlation analysis demonstrated their association with recognized dimensions of banking soundness. K-means and DBSCAN analyses revealed the latent grouping structure of the observations, and Local Outlier Factor analysis enabled the detection of bank-year observations that differed from their local peers. The strong correspondence between the unsupervised classifications and the Altman-based benchmark, together with the superior ROC performance relative to the decision-tree and support vector machine approaches, indicates that anomaly-based unsupervised learning can offer substantial value for early-warning applications. These results are consistent with the growing literature demonstrating the value of advanced machine-learning approaches in bankruptcy and bank-failure prediction [11-13], while also extending the literature by emphasizing that unsupervised techniques may be particularly advantageous when bankruptcy labels are scarce, uncertain, or institutionally influenced.

The present study has several limitations that should be considered when interpreting its findings. First, the quantitative analysis was based on 13 listed banks and 117 bank-year observations, which limits the size and heterogeneity of the sample and may restrict the generalizability of the findings to other banks or financial institutions. Second, the study focused on banks listed on the Tehran Stock Exchange, and therefore unlisted banks, credit institutions, and other financial intermediaries were not included. Third, although the Altman Z-score was used as an external benchmark, this criterion was originally developed for nonfinancial firms and is not specifically designed for banks; therefore, the comparison should be interpreted cautiously. Fourth, the high proportion of observations identified as anomalous suggests that the anomaly threshold and local-density structure may be sensitive to the sample distribution. Fifth, the study relied primarily on annual financial and governance indicators and did not incorporate higher-frequency information, market-based variables, macroeconomic shocks, textual

disclosures, regulatory intervention data, or depositor behavior. Finally, although the qualitative phase strengthened the contextual validity of the indicators, expert judgments may still be influenced by professional experience, institutional background, and subjective interpretation.

Future research should extend the proposed framework using larger and more diverse banking samples across multiple periods and jurisdictions to evaluate the stability and external validity of the findings. Researchers could include private, public, listed, and unlisted banks and examine whether the model performs similarly across different ownership structures and regulatory environments. Future studies could also incorporate quarterly or monthly data to increase the timeliness of bankruptcy-warning signals and evaluate whether anomalies can be detected several periods before severe financial deterioration. The use of alternative unsupervised techniques, including isolation forests, one-class support vector machines, autoencoders, Gaussian mixture models, hierarchical clustering, and deep anomaly-detection architectures, could provide useful comparative evidence. It would also be valuable to integrate macroeconomic indicators, market variables, textual disclosures, news sentiment, interbank exposures, and network measures of systemic risk into the model. Additional research should examine alternative thresholds for Local Outlier Factor and test the sensitivity of the results to different distance metrics, neighborhood sizes, and clustering parameters. Longitudinal validation could further determine whether observations identified as anomalous subsequently experience measurable deterioration in capital, liquidity, asset quality, profitability, or regulatory status.

From a practical perspective, banking regulators and supervisory authorities can use the proposed framework as a supplementary early-warning mechanism for prioritizing institutions that require more intensive monitoring. Banks may also incorporate the model into internal risk-management systems to identify unusual changes in liquidity structure, funding dependence, profitability composition, governance, and management characteristics before these changes develop into severe financial distress. Particular attention should be paid to negative maturity gaps, excessive loan-to-deposit ratios, dependence on non-operating income, and managerial factors because these variables demonstrated substantial discriminatory importance in the present analysis. Supervisors should not interpret an anomaly score as definitive evidence of bankruptcy; rather, it should be treated as a signal requiring further investigation through stress testing, liquidity assessment, asset-quality review, governance evaluation, and examination of funding behavior. The model may also support the development of risk-based supervisory dashboards in which anomaly scores are monitored over time and combined with established prudential indicators. In this way, unsupervised learning can function as a decision-support tool that strengthens preventive supervision, facilitates earlier corrective intervention, and improves the continuous monitoring of bank financial soundness.

Authors' Contributions

Authors equally contributed to this article.

Ethical Considerations

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

Acknowledgments

Authors thank all participants who participate in this study.

Conflict of Interest

The authors report no conflict of interest.

Funding/Financial Support

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

References

[1] C. W. Calomiris and U. Khan, "An Assessment of TARP Assistance to Financial Institutions," *Journal of Economic Perspectives,* vol. 29, pp. 53-80, 2015, doi: 10.1257/jep.29.2.53.

[2] H. Penikas, "History of Banking Regulation as Developed by the Basel Committee on Banking Supervision 1974–2014," *Estabilidad Financiera / Banco de España,* vol. 28, pp. 9-47, 2015.

[3] M. Doumpos, I. Hasan, and F. Pasiouras, "Bank Overall Financial Strength: Islamic versus Conventional Banks," *Economic Modelling,* vol. 64, pp. 513-523, 2017, doi: 10.1016/j.econmod.2017.03.026.

[4] L. Cernohorska, "The Efficiency of Selected Banking Sectors in the European Union," *Financial Assets and Investing,* vol. 8, pp. 5-18, 2017, doi: 10.5817/FAI2017-1-1.

[5] M. P. Tsagkarakis, M. Doumpos, and F. Pasiouras, "Capital Shortfall: A Multicriteria Decision Support System for the Identification of Weak Banks," *Decision Support Systems,* vol. 145, pp. 1-13, 2021, doi: 10.1016/j.dss.2021.113526.

[6] J. Gilanipour, "Examining Factors Affecting Systemic Risk in Iran's Banking Network," (in Persian), *Quarterly Journal of Fiscal and Economic Policies,* vol. 13, no. 50, pp. 6-59, 2025.

[7] M. Morshedzadeh and M. Ghorbani, "Identifying Indicators of Banks Exposed to Bankruptcy Crisis Based on Thematic Analysis," (in Persian), *Quarterly Journal of Investment Knowledge,* vol. 10, no. 40, pp. 1-29, 2021.

[8] B. Mokhtari and S. Khakzad Kahgh, "Examining the Relationship between Social Responsibility, Financial Crisis, and Bankruptcy Probability with Emphasis on Corporate Earnings Quality in the Tehran Stock Exchange," (in Persian), *Accounting and Management Perspective Quarterly,* vol. 2, no. 15, pp. 1-18, 2019.

[9] D. Veganzones and E. Séverin, "An Investigation of Bankruptcy Prediction in Imbalanced Datasets," *Decision Support Systems,* vol. 112, pp. 111-124, 2018, doi: 10.1016/j.dss.2018.06.011.

[10] H. H. Le and J. L. Viviani, "Predicting Bank Failure: An Improvement by Implementing a Machine-Learning Approach to Classical Financial Ratios," *Research in International Business and Finance,* vol. 44, pp. 16-25, 2018, doi: 10.1016/j.ribaf.2017.07.104.

[11] L. Q. Tuan, C. Y. Lin, and H. W. Teng, "Machine Learning Methods for Predicting Failures of US Commercial Bank," *Applied Economics Letters,* vol. 31, pp. 1353-1359, 2023, doi: 10.1080/13504851.2023.2186353.

[12] Z. Rustam, S. Hartini, S. M. N. Islam, F. Novkaniza, F. R. Aszhari, and M. Rifqi, "Enhancing Bankruptcy Prediction of Banks through Advanced Machine Learning Techniques: An Innovative Approach and Analysis," *Procedia Computer Science,* vol. 162, pp. 895-899, 2023.

[13] S. Shetty, M. Musa, and X. Brédart, "Bankruptcy Prediction Using Machine Learning Techniques," *Journal of Risk and Financial Management,* vol. 15, no. 1, pp. 1-10, 2022, doi: 10.3390/jrfm15010035.

[14] A. Tabbakh, "Bankruptcy Prediction Using Robust Machine Learning Model," *Turkish Journal of Computer and Mathematics Education (TURCOMAT),* vol. 12, pp. 3060-3073, 2021, doi: 10.17762/turcomat.v12i2.1955.

[15] F. Mai, S. Tian, C. Lee, and L. Ma, "Deep Learning Models for Bankruptcy Prediction Using Textual Disclosures," *European Journal of Operational Research,* vol. 274, pp. 743-758, 2018, doi: 10.1016/j.ejor.2018.10.024.

[16] F. Mai, S. Tian, C. Lee, and L. Ma, "Deep Learning Models for Bankruptcy Prediction Using Textual Disclosures," *European Journal of Operational Research,* vol. 274, pp. 743-758, 2019, doi: 10.1016/j.ejor.2018.10.024.

[17] A. Karimpour, M. Azhdar, and S. M. Davoudi, "Application of a Deep Learning Algorithm in Identifying Bankruptcy Warning Signs," (in Persian), *Quarterly Journal of Investment Knowledge,* vol. 15, no. 4, pp. 501-524, 2026.

[18] S. Issa, G. Bizel, S. K. Jagannathan, and S. S. C. Gollapalli, "A Comprehensive Approach to Bankruptcy Risk Evaluation in the Financial Industry," *Journal of Risk and Financial Management,* vol. 17, p. 41, 2024, doi: 10.3390/jrfm17010041.

[19] M. Mercadier, A. Tarazi, P. Armand, and J.-P. Lardy, "Monitoring Bank Risk around the World Using Unsupervised Learning," *European Journal of Operational Research,* vol. 324, pp. 590-615, 2025, doi: 10.1016/j.ejor.2025.01.036.

[20] A. Taghvaei, M. Khezrian, and M. Kargari, "Examining the Presentation of a Banking Liquidity Risk Prediction Model Using a Hybrid Machine Learning Approach," presented at the 9th International Conference on New Ideas in Management, Economics, Accounting and Banking, 2024.