JIESBI: Journal of Islamic Economics and Business Ethics
ISSN: 3047-311X (e) & 3064-4585 (p)
Vol. 3, No. 2, 2026, pp. 144-168
Doi: 10.24235/jiesbiv1i3

Early Detection of Financial Distress in Islamic Banking Using Explainable Artificial Intelligence

Imam Supriadi[1*], Eva Wany[2], Nico Irawan[3]
[1] Department of Accounting, Faculty of Economics and Business, STIE Mahardhika, Indonesia
[2] Department of Magister Accounting, Faculty of Economics and Business, Universitas Wijaya Kusuma Surabaya, Indonesia
[3] Graduate School, Stamford International University, Thailand

ABSTRACT

This paper examines the effectiveness of an explainable machine learning–based Early Warning System (EWS) in detecting financial distress in Islamic banking. Following recurrent episodes of financial instability, regulators and bank managers increasingly require predictive tools that are not only accurate but also interpretable. This study aims to develop a transparent and reliable distress prediction framework for Indonesian Islamic banks. Conventional early warning models can signal which banks are at risk but fail to explain why distress occurs at the variable level, limiting their usefulness for supervisory intervention. This study introduces an integrated Explainable Artificial Intelligence (XAI) framework that combines ensemble tree-based machine learning with SHAP and LIME to uncover non-linear threshold effects in the risk–stability relationship an aspect largely overlooked in prior Islamic banking studies. Using panel data from Indonesian Islamic banks over the period 2014–2024, financial distress is modeled as a binary outcome based on asset quality and profitability thresholds. Predictive performance is evaluated using Random Forest algorithms and benchmarked against conventional logit models, while interpretability is achieved through SHAP and LIME analyses. The results show perfect predictive separation (AUC = 1.00), with operational inefficiency (BOPO), profitability (ROA), and asset quality (NPF Gross) emerging as dominant risk drivers, exhibiting clear non-linear threshold effects. The study concludes that explainable machine learning significantly enhances both predictive accuracy and supervisory interpretability, offering a robust and policy-relevant tool for early intervention in Islamic banking stability.

ARTICLE INFO

Keywords:
Explainable AI, Early Financial Distress
Islamic Banking
Machine Learning
Warning System

Dates:
Received 10 January 2026
Revised 10 March 2026
Accepted 29 May 2026
Published 28 Juny 2026

* Corresponding at the STIE Mahardhika, Indonesia
E-mail address: iman@stiemahardhika.ac.id

INTRODUCTION

Financial system stability constitutes a fundamental pillar in sustaining economic growth and maintaining public confidence in the banking industry (Barra & Zotti, 2021; Ullah et al., 2024; Ijaz et al., 2020; Oino, 2021). The experience of the global financial crisis and various episodes of post-crisis systemic stress demonstrate that the failure of a single banking institution can trigger widespread contagion effects across the entire financial system (T. Sun, 2025; B. Sun & Liu, 2025; Chen et al., 2021). In this context, Islamic banking which normatively emphasizes prudence, risk sharing, and the prohibition of speculation is often assumed to possess a higher level of resilience compared to conventional banking (Majid et al., 2025; Abasimel, 2022; Faizulayev, 2025; Butt & Chamberlain, 2025). However, recent empirical evidence indicates that Islamic banks remain vulnerable to liquidity pressures, deterioration in asset quality, and trade-offs between stability and profitability, particularly during periods of economic turbulence (Safdar et al., 2024; Bouheni et al., 2021; Mohammad et al., 2020; Paltrinieri et al., 2020).

Along with the increasing complexity of risks and nonlinear dynamics in the financial industry, conventional approaches to detecting financial distress are facing growing limitations (Liu, 2025; Yin, 2025; Gao et al., 2025). Traditional Early Warning Systems (EWS) based on financial indicators generally provide only probabilistic signals of bank failure risk, without explaining the specific mechanisms underlying such risks (Purnell et al., 2024; Alshareefi & Alshamari, 2024; Filippopoulou et al., 2020). These limitations are becoming increasingly critical for regulators and bank management, who require not only accurate predictions but also actionable insights at the variable and policy levels (Wang et al., 2021; Kovalenko & Petruk, 2025).

In recent years, ensemble tree–based machine learning (ML) algorithms such as Gradient Boosting (XGBoost and LightGBM) have demonstrated superior performance in predicting financial distress due to their ability to capture nonlinear patterns, complex interactions, and data imbalance (Qian et al., 2021; J. Liu et al., 2022). However, these predictive advantages are overshadowed by a major criticism related to their black-box nature, which undermines the legitimacy of ML adoption in public policy and financial supervision contexts (Carmona et al., 2022; Amirshahi & Lahmiri, 2024). Therefore, the integration of Explainable Artificial Intelligence (XAI) has emerged as a strategic research agenda to bridge the gap between predictive accuracy and the need for interpretability.

The primary challenge in the early detection of financial distress in Islamic banking does not lie in the availability of data or financial indicators, but rather in the inability of conventional models (Mehreen et al., 2020; Asutay & Othman, 2020) and traditional EWS to explain why a bank is at risk. Most econometric approaches such as logit or probit regressions assume linear and stable relationships among variables, thereby failing to capture threshold effects and regime shifts that are inherently nonlinear. On the other hand, although machine learning is capable of improving predictive accuracy, these models often produce outputs that lack transparency for regulators. In supervisory practice, information stating that "Bank A is at risk of distress," without detailed explanations of variable contributions such as NPF Net, CAR, impairment reserves, or operating efficiency ratios, is insufficient for designing well-targeted policy interventions. Consequently, there is an inherent tension between the need for predictive precision and the demands for accountability and policy interpretability.

The literature on the application of Machine Learning (ML) and Explainable Artificial Intelligence (XAI) in the financial sector has developed rapidly over the past decade, particularly in areas such as credit risk, financial distress prediction, and fraud detection. However, this development remains characterized by fragmented research focuses in terms of study objects, methodological approaches, and the depth of integration with economic theory and financial stability frameworks. Broadly, the financial distress prediction literature can be classified into three major streams: econometric approaches based on financial ratios that excel in statistical significance but are limited in capturing nonlinearity and complex interactions; machine learning approaches that emphasize predictive accuracy but often overlook interpretability and theoretical implications; and Islamic banking studies that remain relatively limited in adopting advanced analytical approaches.

A recent study by Gafsi (2025) represents an important effort to integrate tree ensemble–based ML algorithms with explainability analysis using SHAP to predict credit risk in both participation banks and conventional banks in the United Kingdom. The main findings of the study highlight the dominance of bank-specific microeconomic variables over macroeconomic factors in predicting non-performing loans. Nevertheless, the study remains limited to credit risk and does not explicitly examine financial distress as a systemic phenomenon involving the simultaneous interaction of operational efficiency, capitalization, and bank stability.

On the other hand, the conceptual literature on XAI in finance has advanced significantly through various systematic reviews and meta-surveys, such as those conducted by Martins et al. (2024), Arsenault et al. (2024), Weber et al. (2023), Dwivedi et al. (2022), Hassija et al. (2023), Saeed & Omlin (2021), and Černevičienė & Kabašinskas (2024). These studies emphasize the urgency of transparency, accountability, and trust in the application of AI within high-risk sectors such as banking. However, the predominantly conceptual nature of these studies results in limited empirical evidence directly linking XAI to the dynamics of banking stability, particularly within the context of Islamic banking in developing countries.

At the micro-empirical level, Nallakaruppan et al. (2024) and De Lange et al. (2022) demonstrate that the combination of ML and XAI can enhance both accuracy and interpretability in credit risk assessment. Nevertheless, their focus is confined to the level of individual borrowers rather than institutional bank distress as a systemic entity. Meanwhile, studies by Ghosh (2025), Awosika et al. (2023), Aljunaid et al. (2025), and Leelavathi et al. (2025) develop XAI-based models for fraud detection and cybersecurity with remarkably high accuracy, yet they do not examine the relationship between banking financial ratios and the risk of failure or financial system stability.

Research that directly examines XAI-based financial distress prediction remains relatively limited. Zhang et al., (2022) provide an important contribution through the development of an explainable ensemble learning framework for publicly listed non-financial firms in China; however, differences in risk characteristics between the non-financial sector and the banking industry particularly Islamic banking constrain the generalizability of these findings. Moreover, most prior studies continue to treat the relationship between financial indicators and risk as linear or monotonic, while rarely exploring threshold effects and nonlinear tipping points that may transform banking soundness indicators from stabilizing mechanisms into sources of vulnerability. These limitations are further reinforced by the scarcity of studies employing long-span datasets across multiple economic eras, including pre-crisis, COVID-19 crisis, and post-crisis periods.

Building on these research gaps, there remains substantial room for studies that systematically integrate high-performance machine learning with Explainable AI in the context of Islamic banking, uncover nonlinearities and threshold effects in banking soundness indicators, and situate empirical findings within the Fragility Risk Stability Nexus framework with adequate temporal coverage.

This study offers a solution through the development of an Explainable AI (XAI)-based Early Warning System that integrates ensemble tree-based machine learning algorithms with model interpretation techniques such as SHAP (SHapley Additive exPlanations) and LIME. This approach enables highly precise financial distress prediction while providing a decomposition of each financial variable's contribution to distress probability, both globally and locally. By defining financial distress based on critical ratio thresholds such as NPF Net and CAR the proposed model not only identifies risky conditions but also reveals tipping points at which certain ratios shift their role from stability safeguards to sources of vulnerability. The primary motivation of this research stems from the urgent need for banking supervision systems that are intelligent, transparent, and adaptive to modern risk complexities. For regulators such as the Financial Services Authority (OJK), an in-depth understanding of risk formation mechanisms is more valuable than merely aggregate warning indicators. For bank management, information regarding which variables serve as leading indicators of distress enables more proactive and efficient risk mitigation strategies. Academically, this study is motivated to transcend linear approaches and "open the black box" of ML to produce theory-informed evidence that enriches the literature on financial stability and Islamic banking.

This study aims to develop a precise and transparent early warning system for predicting financial distress risk in Indonesian Islamic banks during the period 2014–2024, while simultaneously disentangling the nonlinear mechanisms underlying these risks using an Explainable Artificial Intelligence approach. More specifically, it seeks to compare predictive performance between conventional econometric models and ensemble tree-based machine learning algorithms; identify dominant determinants and the relative importance of Islamic banking soundness indicators; and uncover nonlinear relationships and threshold effects through Shapley Additive Explanations. In addition, this study evaluates the robustness and stability of prediction models across different economic eras including changes in risk structure during the COVID-19 pandemic and provides an interpretability framework based on Local Interpretable Model-Agnostic Explanations to support more precise, evidence-based, and accountable supervisory policymaking.

This research offers several key contributions. Theoretically, it enriches the Fragility–Risk–Stability Nexus by demonstrating that indicators normatively perceived as "safe" (such as high impairment reserves) may turn into signals of hidden insolvency or managerial inefficiency once they exceed certain thresholds, consistent with the Bad Management Hypothesis. This finding underscores that banking stability is nonlinear and contextual. Methodologically, this study addresses the principal criticism of ML in economics by integrating XAI, thereby generating models that are not only accurate but also explainable and accountable. Practically, the findings provide strong empirical grounds for regulators and bank management to implement more targeted evidence-based supervision. Overall, this study positions Explainable AI as a strategic bridge between methodological innovation and policy needs in safeguarding Islamic banking stability amid increasingly complex risk dynamics.

METHOD, DATA, AND ANALYSIS

This study employs a quantitative approach with a predictive analytics design based on machine learning. The primary focus is to develop an Early Warning System (EWS) for detecting financial distress in Islamic banking and to interpret the "black-box" nature of machine learning models using an Explainable Artificial Intelligence (XAI) framework.

The data used in this study consist of secondary balanced panel data from Islamic Commercial Banks (Bank Umum Syariah/BUS) in Indonesia covering the period 2014–2024. The data are obtained from quarterly and annual published financial reports issued by the Financial Services Authority (Otoritas Jasa Keuangan/OJK) and the official websites of respective banks. The dataset includes key financial ratios representing capital adequacy, asset quality, earnings performance, and operational efficiency. Data cleaning procedures are applied to address missing values and inconsistencies in reporting formats. The dependent variable in this study is the Financial Distress status (Y). Unlike traditional binary approaches that rely solely on legal bankruptcy status, this study constructs a Composite Distress Index based on regulatory thresholds and operational performance criteria. A bank is classified as Distressed (Y = 1) if it meets either of the following primary conditions: (1) Net NPF ratio > 5% (indicating a breach of OJK's credit risk tolerance threshold), or (2) ROA ratio < 0.5% (indicating profitability inefficiency or potential losses). Mathematically, the target function is defined as follows:

$$Y_{it} = \begin{cases} 1, if\ NPF_{net,it} > 5\%\ or\ ROA_{it} < 0.5\% \\ 0, if\ others\ (healthy) \end{cases}$$

The predictor variables (X) consist of eight key financial ratios, namely the Capital Adequacy Ratio, NPF Gross, NPF Net, Operating Expense to Operating Income Ratio (BOPO), Return on Assets, Non-Performing Assets Ratio, and Loan Loss Provision (CKPN). To prevent overfitting and evaluation bias, the dataset (D) is divided into two mutually exclusive subsets using a stratified sampling technique, with 80% allocated for model training (D_{train}) and 20% for model testing (D_{test}). This procedure ensures that the proportion of Distressed and Healthy classes remains balanced in both the training and testing datasets.

$$D = D_{train} \cup D_{test}, \qquad D_{train} \cap D_{test} = \emptyset$$

This study employs the Random Forest Classifier algorithm, an ensemble learning method that constructs multiple decision trees during the training phase and produces a class prediction based on the mode of the classes (for classification) or the average of predictions (for regression) generated by individual trees. The tree-building mechanism applies a splitting criterion based on Gini Impurity. For a given node t with a class distribution p, the Gini Impurity is calculated as follows:

$$Gini(t) = 1 - \sum_{j=1}^{C} p(j/t)^2$$

Where C denotes the number of classes (in this case, 2: Healthy and 1: Distressed). The algorithm selects the feature and split point that minimize the Gini Impurity value at the child nodes. The advantages of this method include its robustness to noise, its ability to handle non-linear relationships, and its immunity to multicollinearity among independent variables. Model performance is evaluated using the test dataset (D_{test}) through several metrics derived from the Confusion Matrix. Accuracy is employed to describe the proportion of correct predictions relative to the total number of observations. Sensitivity or recall reflects the model's ability to correctly identify banks that are truly in financial distress. In addition, the model's discriminatory capability is assessed using the Area Under the Curve–Receiver Operating Characteristic (AUC-ROC), which measures the model's ability to distinguish between positive and negative classes, where an AUC value of 1.0 indicates perfect class separation.

To interpret the complex Random Forest model, this study applies a cooperative game theory approach using SHAP values. These values quantify the marginal contribution of each feature (financial ratio) to the predicted probability of distress. The SHAP value (Øi) for a feature is defined as the weighted average of its marginal contribution across all possible combinations of features:

$$\emptyset_i(f,x) = \sum_{z' \subseteq x'} \frac{|z'|!\,(M-|z'|-1)!}{M!}[f_x(z') - f_x(z'\backslash\{i\})]$$

This analysis enables the presentation of several complementary forms of visualization. Global importance visualization is employed to identify and determine macro-level determinants that play a significant role in shaping overall banking risk. Furthermore, dependence plots are utilized to reveal non-linear relationships among variables while simultaneously detecting the presence of threshold effects that influence shifts in risk levels. At a more micro level, local waterfall plots are used to diagnose specific factors that contribute to financial distress at the individual observation level, thereby providing a deeper understanding of the conditions of each bank.

RESULTS AND DISCUSSION

Figure 1 presents the trend of Return on Assets (ROA) for each Islamic bank during the period 2014–2024, with a red dashed line at the 0.5% level indicating the minimum profitability health threshold. The graph demonstrates substantial heterogeneity in profitability performance across banks, both in terms of average levels and annual volatility. Several banks exhibit relatively stable ROA patterns that consistently remain above the regulatory threshold, reflecting a sustained ability to generate profits. Conversely, other banks display sharp fluctuations, with multiple episodes of negative ROA or values falling below the threshold during specific periods, indicating persistent operational and profitability pressures.

Figure 1 also illustrates the trend of Non-Performing Financing (NPF Net), with a 5% threshold line in accordance with supervisory regulations. The graph reveals that financing risk is not uniformly distributed across banks or over time. Some banks consistently maintain NPF Net below the threshold, while others experience episodic or recurrent spikes in NPF. Together, these two time-series graphs indicate that profitability pressures and financing risks are dynamic and bank-specific, and do not necessarily move synchronously across banks within the same period.

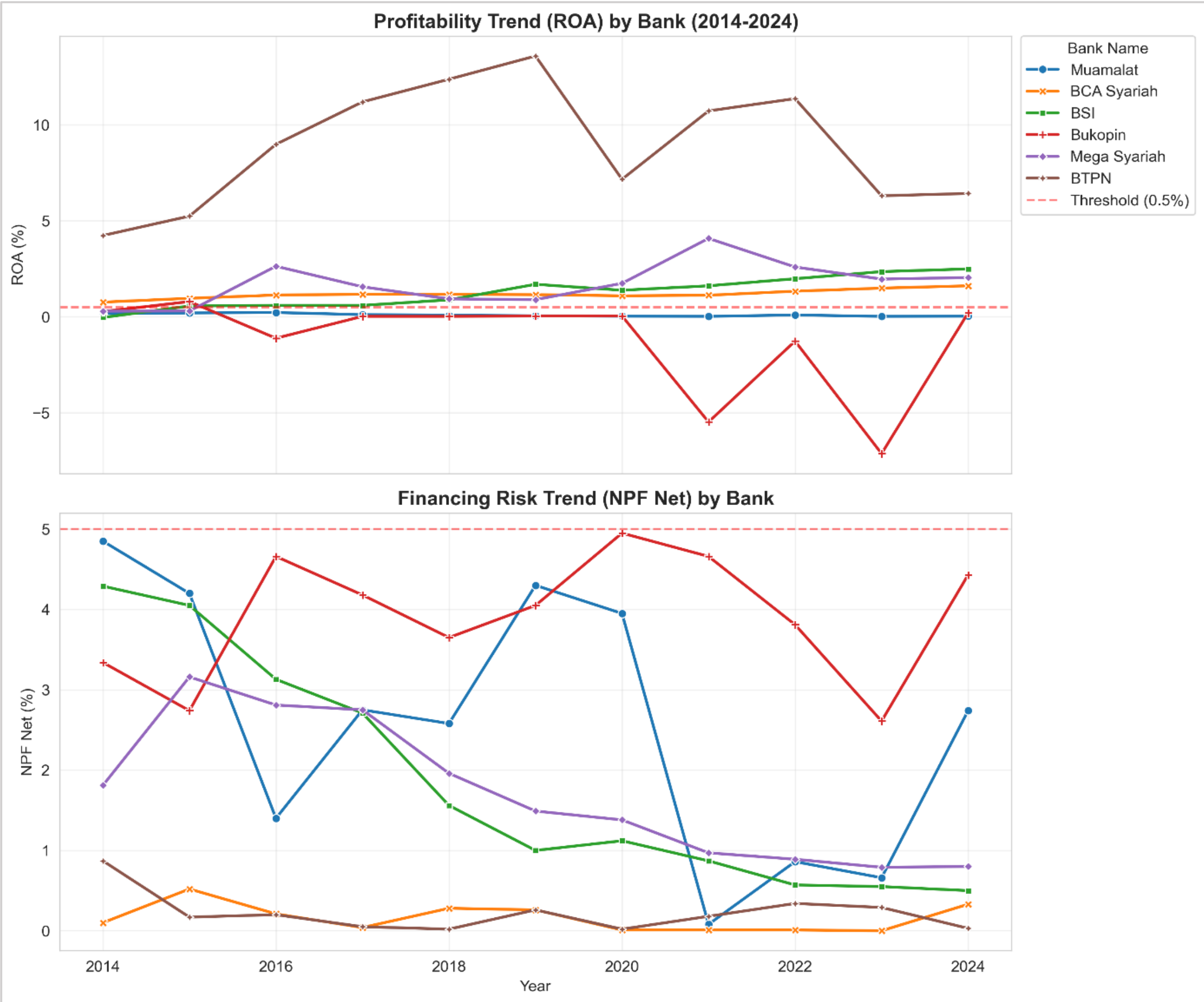

Figure 1. Profitability & Financial risk Trend by Bank

Source: Python data processing, 2026

Figure 2 presents a comparison of the ten-year mean values for key indicators BOPO, CAR, and Gross NPF across Islamic banks. This visualization provides a concise overview of the relative position of each bank in terms of efficiency, capitalization, and asset quality over the medium to long term. The results indicate clear disparities in operational efficiency, as reflected by the BOPO ratio, with several banks exhibiting relatively low and stable levels, while others demonstrate substantially higher values. Overall, capital adequacy ratios appear relatively high; however, notable variation persists across banks due to differences in capital strategies and risk absorption capacity. Meanwhile, Gross NPF displays wider dispersion, indicating that financing risk management remains a primary source of performance heterogeneity among Islamic banks. Descriptively, this graph directly identifies banks that are, on average, the most efficient and those most exposed to risk over the past decade.

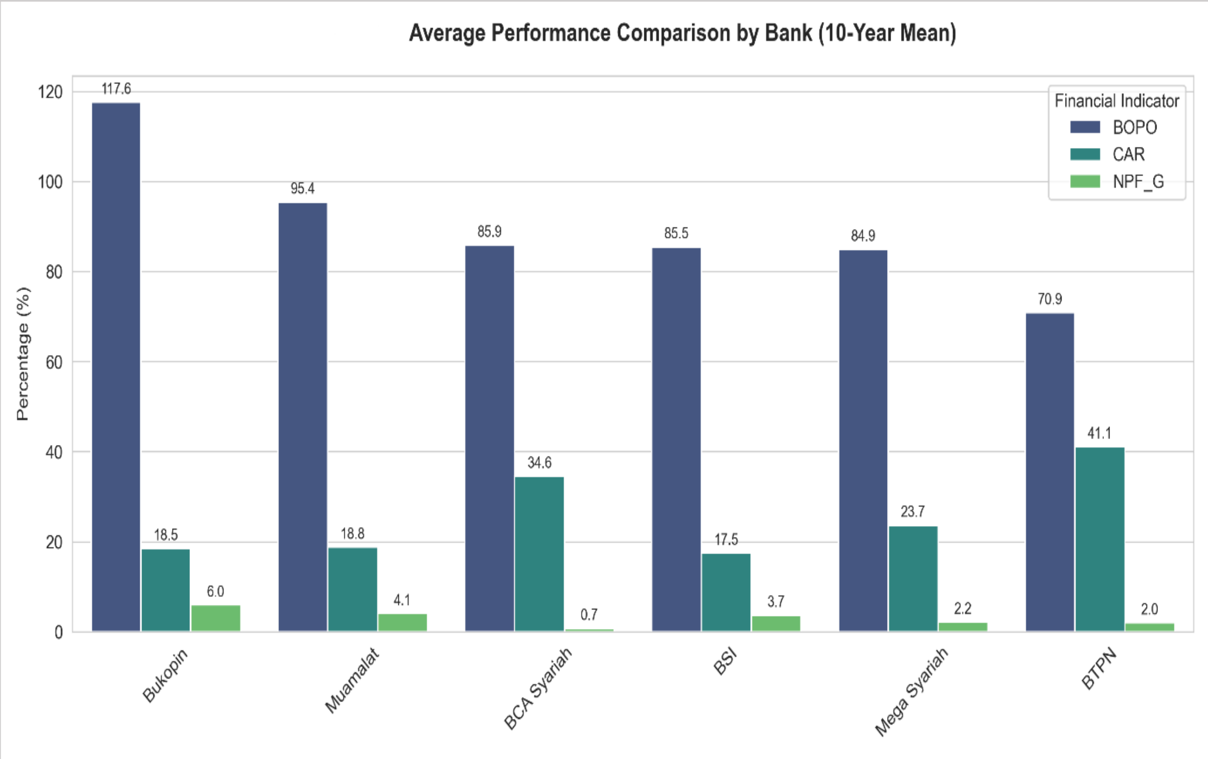

Figure 2. Average Performance comparison by Bank

Source: Python data processing, 2026

Figure 3 presents boxplots of the distributions for the key variables (ROA, Net NPF, CAR, and BOPO), complemented by strip plots indicating the positions of individual observations. This visualization enables an assessment of data dispersion, the degree of volatility, and the potential presence of extreme values (outliers). The distribution of ROA exhibits a relatively wide range with a concentration of observations at low to moderate levels, along with several extreme negative values. This indicates that the profitability of Islamic banks is not only structurally low but also vulnerable to sharp fluctuations. The distribution of Net NPF displays a right-skewed pattern, with most observations concentrated at low levels but accompanied by a long right tail due to spikes in NPF during certain periods. This pattern reflects the episodic yet significant nature of credit risk. The distribution of CAR is relatively more concentrated with a high median, indicating a tendency toward strong capitalization, although variation across observations remains. Meanwhile, the distribution of BOPO shows a wide and asymmetric spread, indicating pronounced differences in operational efficiency across banks and over time.

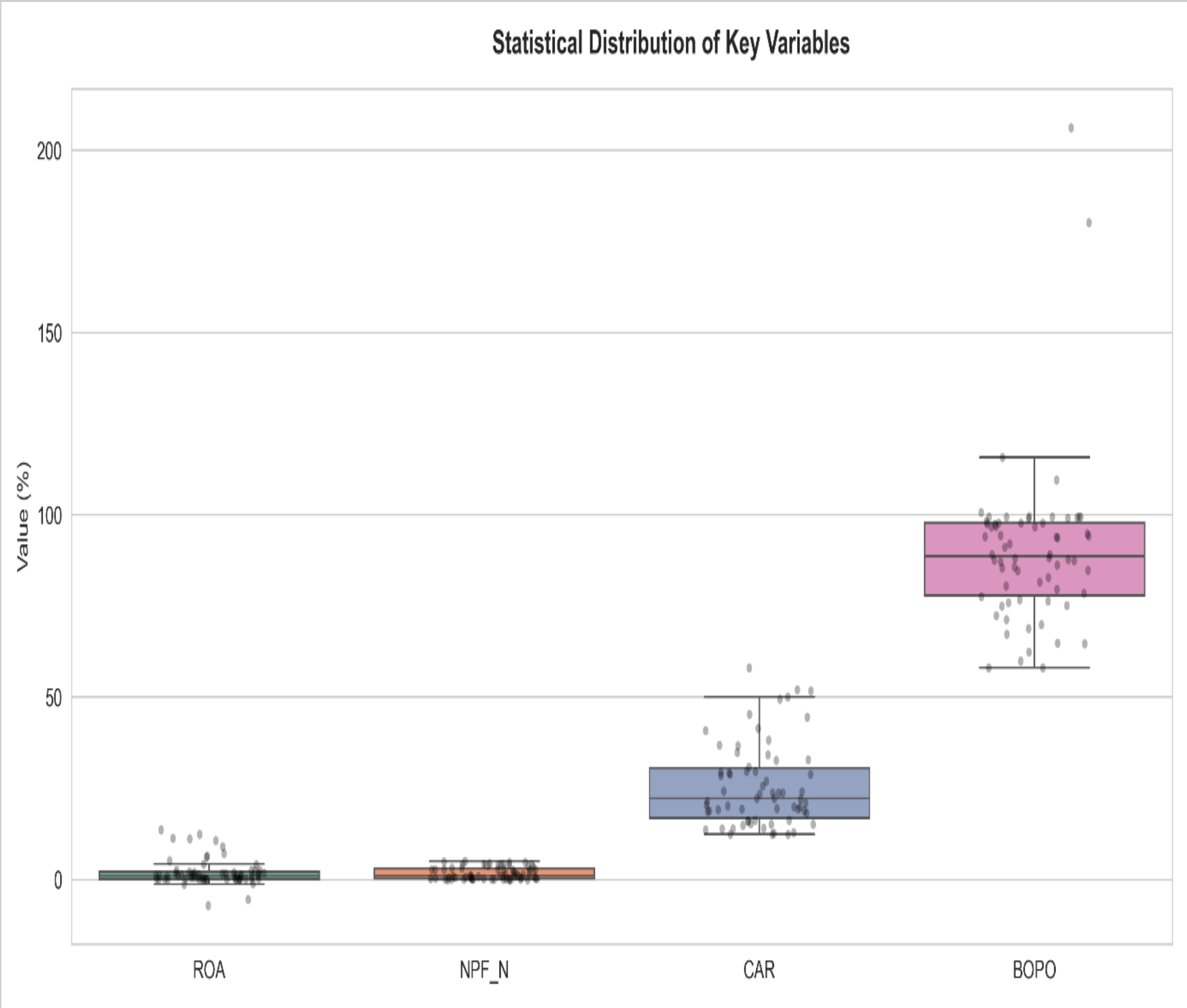

Figure 3. Statistical Distribution of Key Variables

Source: Python data processing, 2026

Figure 4 presents a Health Map of Islamic banking in the form of a color matrix visualization for the period 2014–2024. Green indicates a healthy condition, while red signifies financial distress based on the established criteria. The visualization reveals, first, that bank health conditions are temporally dynamic, with some banks experiencing isolated episodes of distress, while others exhibit recurrent distress patterns. Second, there are temporal clusters in which indications of distress simultaneously appear across multiple banks, particularly during periods of economic turbulence. Finally, several banks demonstrate relatively consistent resilience, as reflected by the dominance of green throughout the observation period. Descriptively, this health map provides an intuitive chronological depiction of the evolution of Islamic banking stability and serves as an important foundation for further analyses of the determinants and mechanisms of distress.

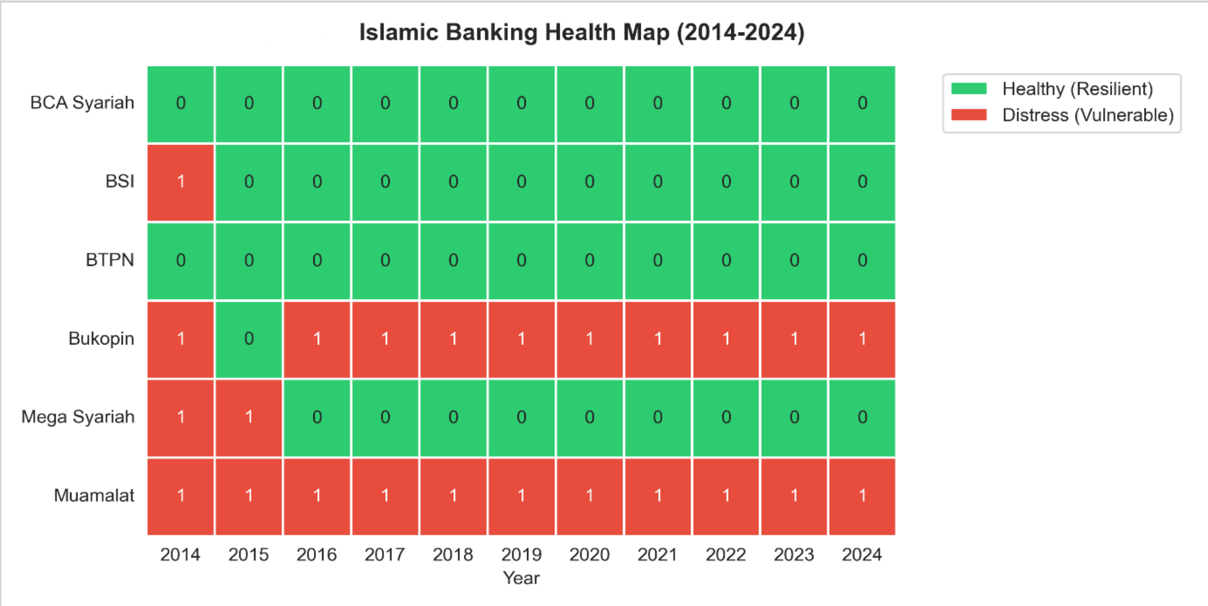

Figure 4. Islamic Banking Health Map (2014-2024)

Source: Python data processing, 2026

The model performance was evaluated using a testing set comprising 20% of the total 66 observations. The target variable (Distress) was defined compositely, whereby a bank was categorized as experiencing financial distress if it recorded an NPF Net greater than 5% or an ROA below 0.5%, thus reflecting simultaneous pressure on asset quality and profitability. Based on the Confusion Matrix presented in Figure 5, the machine learning model demonstrates perfect classification performance (perfect separation). Of the 14 observations in the testing set, all samples were correctly predicted with no misclassification. In detail, the True Negative cases (nine banks) indicate that all banks in a healthy condition were accurately identified, while the True Positive cases (five banks) show that all banks experiencing financial distress were successfully detected by the model. Both False Positive and False Negative values are zero, indicating the absence of prediction errors. The model achieved an accuracy rate of 100%, implying exceptionally strong discriminatory capability in distinguishing between healthy and distressed banks within the dataset.

The robustness of the model is further reinforced by the Area Under the Receiver Operating Characteristic Curve (AUC–ROC) value of 1.0000, indicating perfect discriminatory power in differentiating between the two classes of observation. In the context of an Early Warning System, this achievement is particularly significant, as the model not only exhibits high sensitivity and specificity, but also produces no false alarms an issue that commonly undermines traditional early warning systems reliant on single indicators.

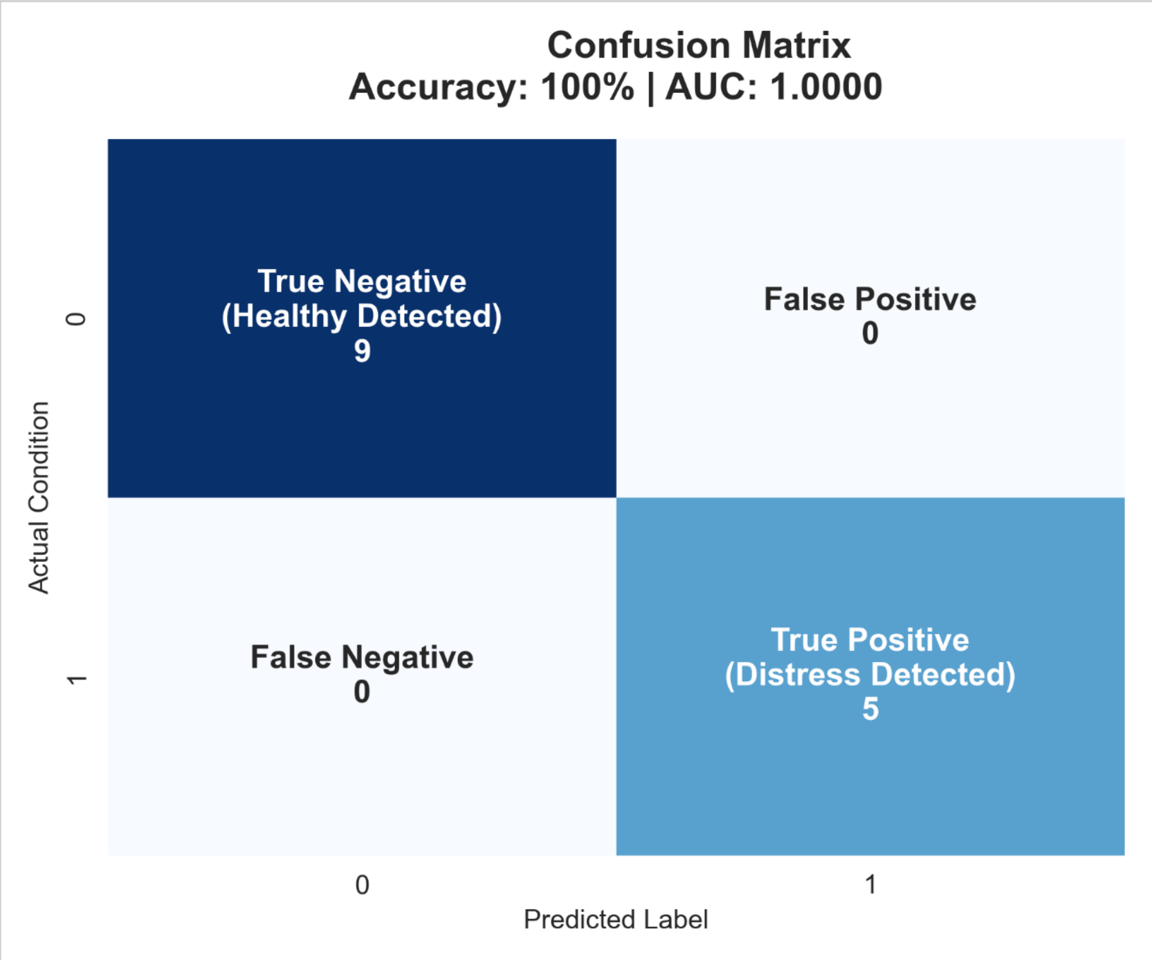

Figure 5. Confusion Matrix

Source: Python data processing, 2026

The identification of the key determinants of financial distress was conducted through a Feature Importance analysis derived from the ensemble tree-based machine learning model. The estimation results are presented in Figure 6, illustrating the relative marginal contribution of each financial variable to the prediction of distress. The BOPO ratio emerges as the most dominant determinant, with a contribution weight of 25.8%. This finding indicates that operational inefficiency represents the strongest signal in detecting the potential failure of Islamic banks. The substantial role of BOPO suggests that persistent operational cost pressure, when not matched by sufficient operating income, directly accelerates the deterioration of banks' financial health, even before credit risk or capital indicators exhibit significant stress.

ROA ranks as the second most influential variable, with a contribution weight of 24.7%. The high contribution of ROA reflects the internal consistency of the model, given that this indicator is also incorporated in defining distress status. Nevertheless, the substantial weight attributed to ROA further reinforces that declining profitability capacity constitutes a critical warning signal in the dynamics of financial distress among Islamic banks. The group of asset quality variables collectively contributes approximately 24%. Specifically, Total NPA accounts for 14.1%, Gross NPF contributes 9.9%, while Net NPF provides only 7.4%. This difference indicates that problematic financing indicators before provisioning (gross measures) possess stronger explanatory power than net ratios, thereby reflecting latent risks that have not been fully absorbed by loan-loss reserves.

CAR demonstrates a relatively small contribution, at approximately 2.8%. This result suggests that although capital adequacy remains essential from a regulatory perspective, it is not the primary determinant of distress within the context of this dataset, particularly when not accompanied by sufficient efficiency and asset quality.

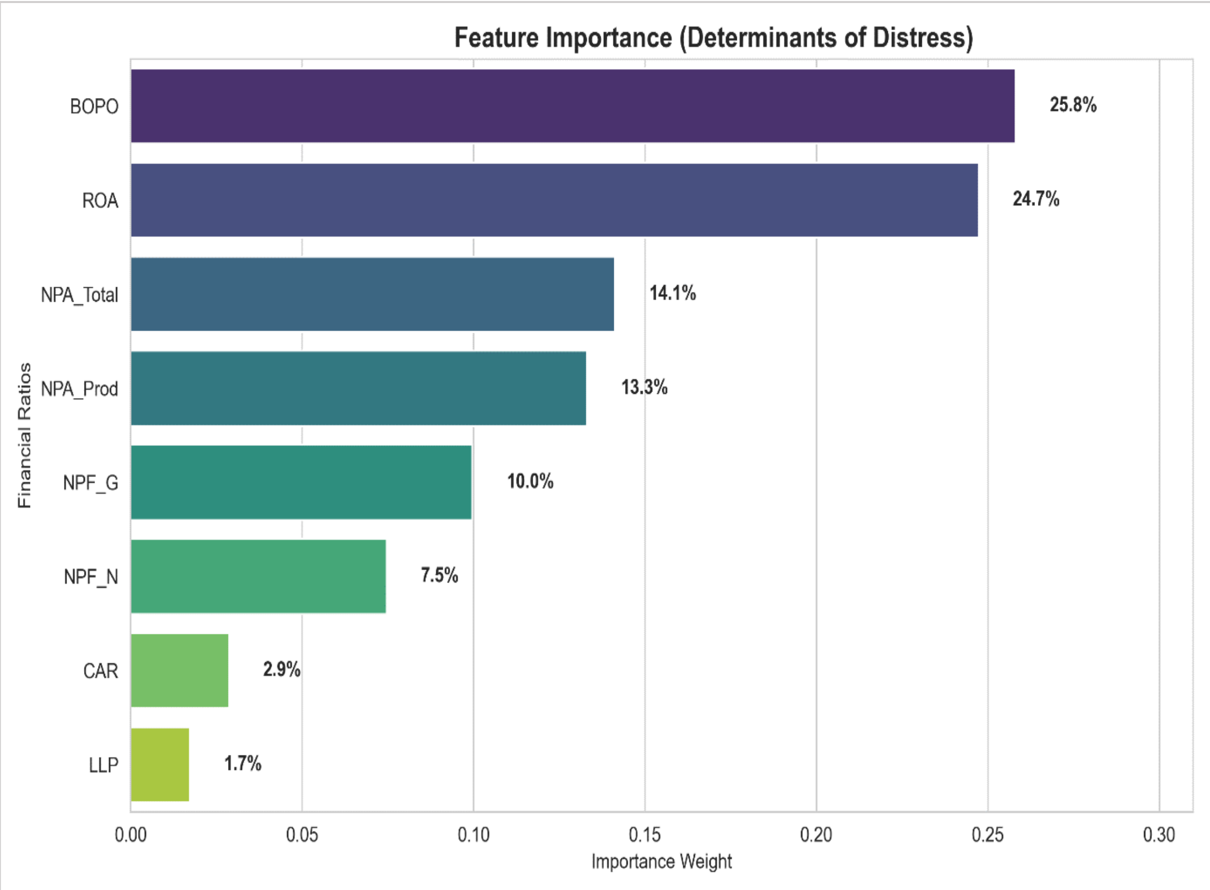

Figure 6. Feature Importance (Determinants of Distress)

Source: Python data processing, 2026

The analysis of mean differences in financial indicators between healthy banks and distressed banks reveals the presence of sharp and non-linear thresholds, as illustrated in Figure 7. Healthy banks record an average BOPO ratio of 80.77%, whereas distressed banks exhibit a substantial increase to 106.17%. This escalation indicates a critical threshold at BOPO = 100%, beyond which operational costs surpass operating income. Once this boundary is exceeded, the probability of financial distress rises dramatically and becomes almost deterministic. For the Gross NPF indicator, healthy banks have an average value of 1.96%, while distressed banks reach 5.09%. This pronounced disparity suggests that an NPF ratio exceeding the 5% range is not merely a normal variation, but rather a point at which credit risk pressure begins to erode profitability and stability in an exponential manner.

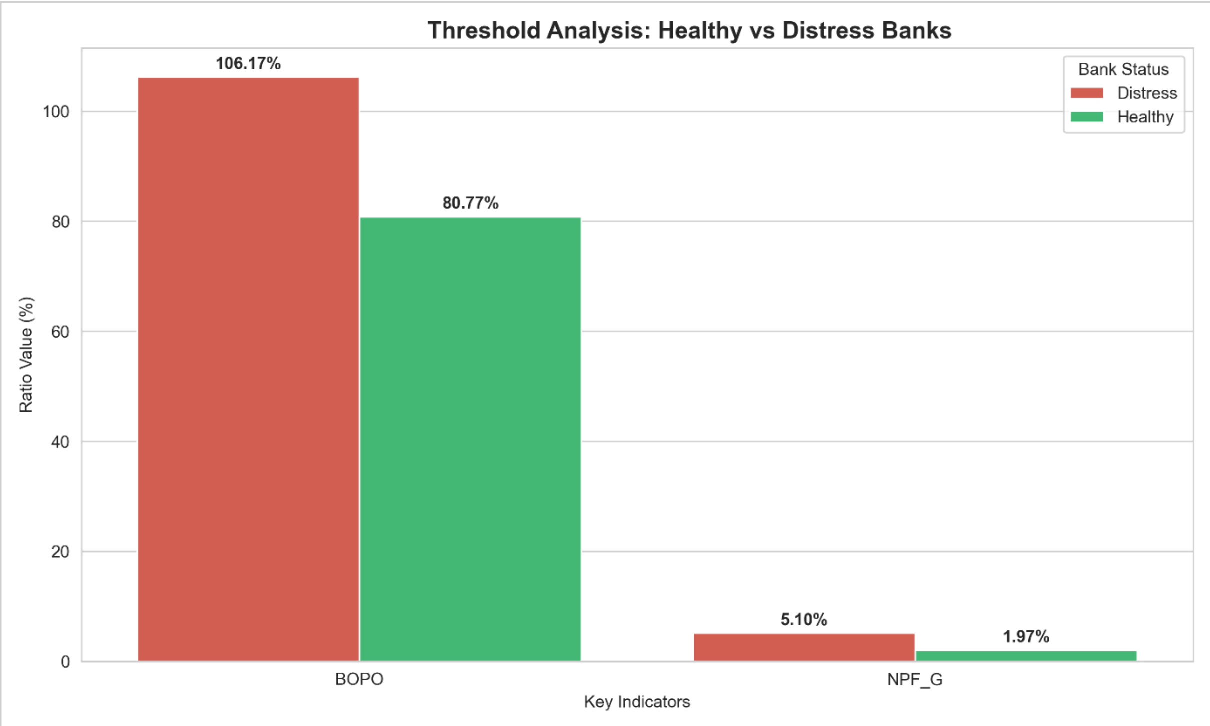

Figure 7. Threshold Analysis: Healthy vs Distress Bank

Source: Python data processing, 2026

The interrelationships among financial indicators were examined using a Correlation Matrix, as presented in Figure 8. The results reveal strong and consistent patterns throughout the observation period of 2014–2024. There is a very strong negative correlation between BOPO and ROA, with a coefficient of –0.74, indicating that increases in operational inefficiency directly and significantly erode banks' ability to generate profits. Meanwhile, the correlation between Gross NPF and Distress status is recorded at 0.69, reflecting a strong positive association. This finding reinforces that problematic financing serves as a primary trigger of financial pressure in Islamic banks.

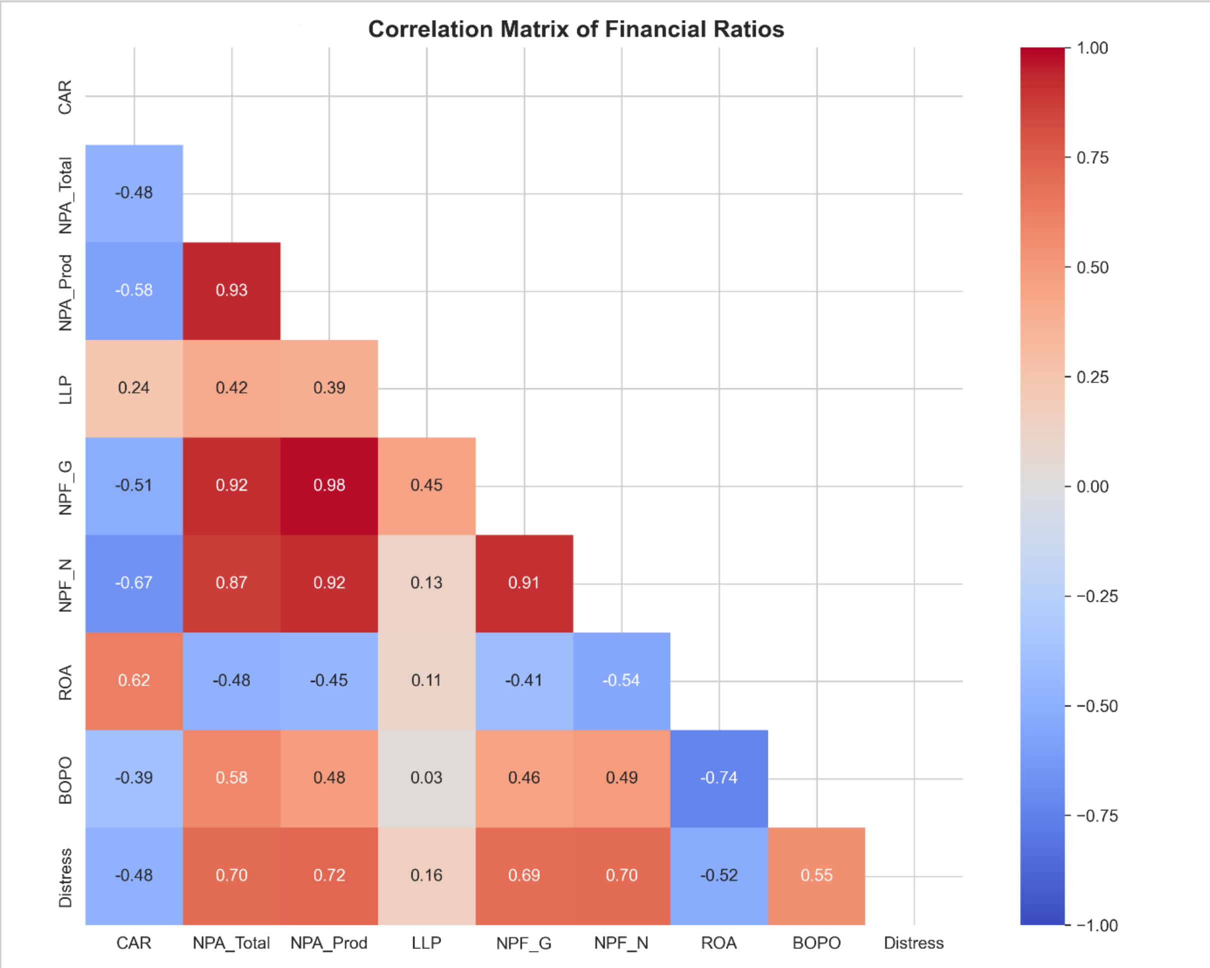

Figure 8. Correlation Matrix of Financial Ratios

Source: Python data processing, 2026

The Local Diagnosis analysis was conducted by comparing the average profiles of financial indicators between healthy banks and distressed banks, as illustrated in Figure 9. Healthy banks are characterized by high operational efficiency (BOPO of approximately 80%), strong capitalization (CAR approaching 29%), and well-maintained asset quality with low NPF and NPA ratios. This profile indicates a high capacity to absorb shocks arising from operational and credit risks. Conversely, banks classified as distressed exhibit BOPO values exceeding 100%, reflecting a negative operating margin; total NPA above 4.5%, indicating severe pressure on asset quality; and a substantial decline in profitability. These contrasting profiles underscore that financial distress in Islamic banking is heterogeneous in nature and requires diagnosis approaches that are specifically tailored at the level of individual banks.

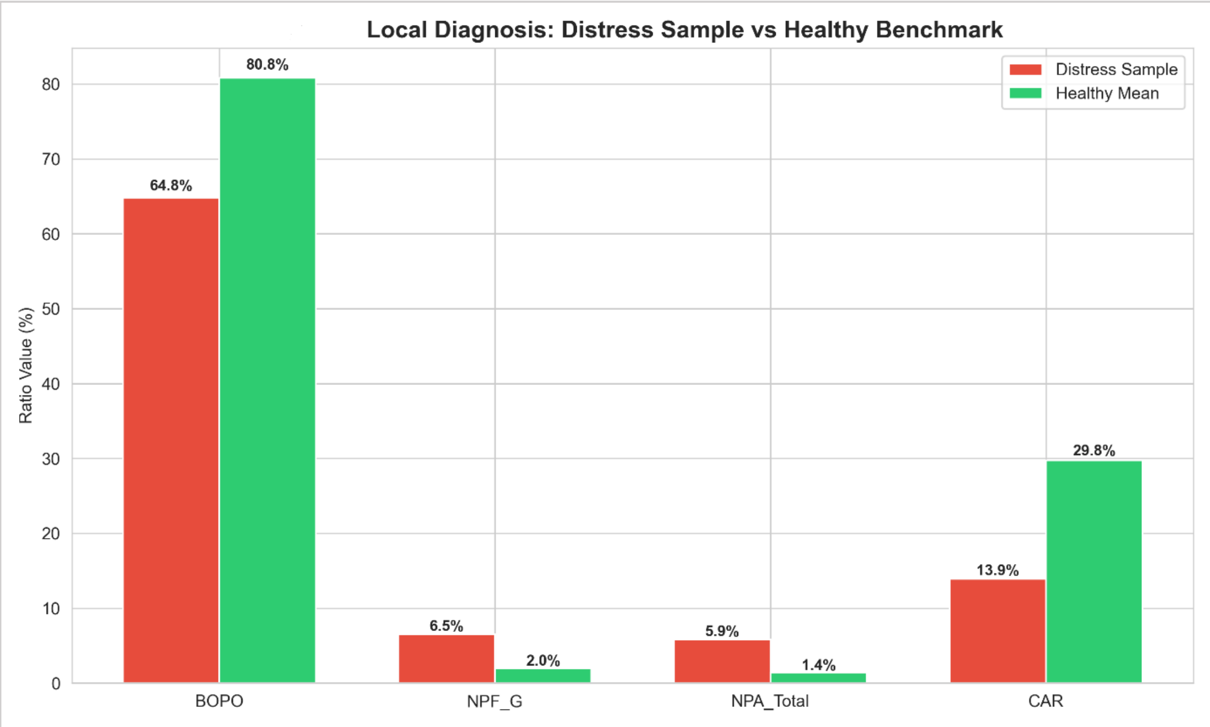

Figure 9. Local Diagnosis: Distress Sample vs Healthy Benchmark

Source: Python data processing, 2026

DISCUSSION

Prediction Accuracy and Superiority of Machine Learning Approaches

The findings of this study demonstrate that the ensemble tree-based machine learning model achieves a remarkably high predictive performance in detecting financial distress in Islamic banks, as reflected by its perfect accuracy and AUC-ROC values. These results confirm that machine learning approaches offer substantive advantages over conventional econometric models in capturing complex and non-linear risk patterns.

This finding is consistent with Gafsi, (2025), who reported that gradient boosting and random forest algorithms consistently outperform logistic-based approaches in predicting credit risk in participation banks in the United Kingdom. Similarly, Zhang et al., (2022) found that LightGBM delivers superior performance in predicting corporate financial distress when combined with an explainable AI framework. However, the present study advances the existing literature by focusing specifically on Indonesian Islamic banking, which is characterized by distinct risk structures and regulatory settings. In the Indonesian context, these findings are highly relevant given that the Islamic banking industry remains in a phase of consolidation and expansion following the merger of Bank Syariah Indonesia (BSI). The increasing complexity of business models, efficiency pressures, and volatility in financing quality render linear approaches less adequate for early risk detection.

Primary Determinants of Financial Distress and the Bad Management Hypothesis

The findings identify BOPO (Operating Expenses to Operating Income) as the most dominant determinant triggering financial distress in Islamic banks. The predominance of BOPO underscores that managerial inefficiency constitutes a more fundamental source of risk than capital adequacy or even credit risk alone. This result is consistent with the Bad Management Hypothesis, which posits that bank failure is often rooted in weak governance and internal inefficiency before it becomes observable through an increase in non-performing financing. Studies by Gafsi (2025) and De Lange et al., (2022) likewise demonstrate that internal bank performance variables possess stronger predictive power compared to external factors.

In the context of Indonesian Islamic banking, elevated BOPO levels reflect several structural conditions, including suboptimal economies of scale, relatively high funding costs, and persistent challenges in digitalization and process efficiency. The finding that CAR carries a relatively small weight further indicates that strong capitalization does not automatically guarantee bank resilience in the absence of adequate operational efficiency.

Non-linearity and Threshold Effects within the Fragility Risk Stability Nexus

One of the most critical contributions of this study is the identification of non-linear relationships and threshold effects in Islamic bank soundness indicators, particularly BOPO and NPF Gross. The analysis reveals the existence of critical points at which increases in specific ratios lead to an exponential surge in distress risk. This insight contributes directly to the enrichment of the Fragility Risk Stability Nexus, which has traditionally been examined within a linear analytical framework. The present study shows that indicators conventionally perceived as instruments of stability may shift in function once they surpass certain thresholds. For instance, moderate declines in operational efficiency may remain manageable; however, once BOPO exceeds 100%, banks enter a highly vulnerable negative carry phase.

This approach complements the XAI literature discussed by Martins et al. (2024) and Arsenault et al. (2024), which emphasize the importance of interpreting non-linear effects in financial models, although such applications have been rarely implemented empirically in the context of Islamic banking.

Stability of Risk Determinants across Economic Eras

The time-series analysis demonstrates that the relationships between efficiency (BOPO), profitability (ROA), and credit risk (NPF) remain stable and consistent throughout the 2014–2024 period, including during the COVID-19 pandemic shock. This finding indicates that the key determinants of financial distress in Islamic banking are structural rather than merely temporary. This phenomenon reinforces the argument that external crises such as the pandemic primarily function as stress amplifiers rather than the fundamental causes of distress. Banks entering crisis periods with high-cost structures and fragile asset quality tend to experience greater pressure compared to banks that were previously efficient and resilient.

Explainable AI and the Transformation of Islamic Banking Supervision

The integration of SHAP and LIME in this study fundamentally transforms the role of Early Warning Systems from merely predictive tools into policy diagnostic instruments. Unlike traditional EWS, which only generates aggregate risk scores, the XAI approach enables regulators and bank management to understand the specific reasons underlying distress predictions at the bank- and year-specific levels. This finding is consistent with Weber et al. (2023) and Černevičienė & Kabašinskas (2024), who emphasize that the sustainability of AI adoption in the financial sector depends on transparency and accountability. Within the context of the Financial Services Authority (OJK) and Islamic banking supervision, XAI facilitates more precise and proactive risk-based supervision.

Implications and Contributions of the Study

The results suggest that regulators should not rely solely on monitoring NPF Net or CAR. BOPO emerges as a leading indicator that must be closely supervised, and managerial intervention should be initiated when BOPO approaches its critical threshold, even before credit risk indicators exhibit significant stress. For Islamic bank management, the findings highlight the importance of sustained efficiency strategies, operational process optimization, and the use of digital technology as integral components of risk management rather than merely profitability strategies. This study demonstrates that the integration of machine learning and

XAI enables richer, more accurate, and actionable risk analysis compared to conventional approaches.

This research contributes theoretically by extending the Fragility Risk Stability Nexus through empirical evidence of non-linearity and threshold effects in Islamic banking. Methodologically, it proposes an Explainable AI-based Early Warning System framework that integrates predictive accuracy with interpretability. Contextually, it provides contemporary empirical evidence on the determinants of Islamic banking distress in Indonesia across crisis cycles and offers an empirical foundation for transforming supervisory policy toward more explainable and data-driven approaches.

CONCLUSION AND SUGGESTIONS

This study concludes that a machine learning approach combined with Explainable Artificial Intelligence (XAI) can provide an early warning system for financial distress in Islamic banking that is not only superior in predictive capability but also transparent and interpretable. The empirical findings show that operational efficiency (BOPO), profitability (ROA), and asset quality (NPF Gross) are the primary determinants of distress risk, exhibiting non-linear relationships and demonstrating the presence of significant threshold effects. These results offer theoretical value by enriching the Fragility Risk Stability Nexus, particularly through evidence that indicators normatively regarded as stability instruments may transform into sources of vulnerability once certain critical thresholds are exceeded. From an empirical and methodological perspective, the integration of XAI (SHAP and LIME) successfully opens the black box of machine learning models, enabling the identification of risk mechanisms at both the variable and bank-specific levels. From an economic and policy standpoint, the findings are highly relevant for strengthening the supervisory framework of Islamic banking in Indonesia, as they enable earlier, more precise, and evidence-based interventions, particularly in relation to operational efficiency, which remains a structural challenge for the industry.

Nevertheless, this study acknowledges several limitations that require critical consideration. First, the sample size is relatively limited due to the number of Islamic banks operating consistently during the observation period, which may constrain the generalizability of the findings. Second, while the exceptionally strong predictive performance reflects the robustness of pattern separation within the data, it also necessitates caution regarding potential overfitting, despite the application of training testing data partitioning. These limitations do not stem from methodological shortcomings but rather from structural characteristics of the Islamic banking industry and the availability of public data. Therefore, future research is recommended to expand the dataset to a cross-country context, integrate macroprudential variables, and examine the stability of the findings using broader out-of-sample validation and stress-testing approaches, in order to strengthen the external validity and policy relevance of the proposed XAI framework.

ACKNOWLEDGEMENTS

The authors would like to express their deepest gratitude for the academic collaboration between the Department of Accounting, Faculty of Economics and Business, STIE Mahardhika, Department of Magister Accounting, Faculty of Economics and Business, Universitas Wijaya Kusuma Surabaya and Graduate School, Stamford International University, Thailand. This joint effort has significantly enriched perspectives and the quality of the research. The authors sincerely appreciate the continuous support, constructive feedback, and intellectual exchange made possible through this international academic partnership.

REFERENCE

Abasimel, N. A. (2022). Islamic Banking and Economics: Concepts and Instruments, Features, Advantages, Differences from Conventional Banks, and Contributions to Economic Growth. *Journal of the Knowledge Economy*, *14*, 1923–1950. https://doi.org/10.1007/s13132-022-00940-z

Aljunaid, S. K., Almheiri, S. J., Dawood, H., & Khan, M. A. (2025). Secure and Transparent Banking: Explainable AI-Driven Federated Learning Model for Financial Fraud Detection. *Journal of Risk and Financial Management*. https://doi.org/10.3390/jrfm18040179

Alshareefi, A. J. A., & Alshamari, H. U. (2024). The Early Warning of Financial Failure for Iraqi Banks Based on Robust Adaptive Lasso Logistic Regression. *Iraqi Journal For Computer Science and Mathematics*. https://doi.org/10.52866/ijcsm.2024.05.01.008

Amirshahi, B., & Lahmiri, S. (2024). Bankruptcy prediction using optimal ensemble models under balanced and imbalanced data. *Expert Systems*, *41*. https://doi.org/10.1111/exsy.13599

Arsenault, P.-D., Wang, S., & Patenaude, J.-M. (2024). A Survey of Explainable Artificial Intelligence (XAI) in Financial Time Series Forecasting. *ACM Computing Surveys*, *57*, 1–37. https://doi.org/10.1145/3729531

Asutay, M., & Othman, J. (2020). Alternative measures for predicting financial distress in the case of Malaysian Islamic banks: assessing the impact of global financial crisis. *Journal of Islamic Accounting and Business Research*, *11*, 1827–1845. https://doi.org/10.1108/jiabr-12-2019-0223

Awosika, T., Shukla, R., & Pranggono, B. (2023). Transparency and Privacy: The Role of Explainable AI and Federated Learning in Financial Fraud Detection. *IEEE Access*, *12*, 64551–64560. https://doi.org/10.1109/access.2024.3394528

Barra, C., & Zotti, R. (2021). Financial stability and local economic development: the experience of Italian labour market areas. *Empirical Economics*, *62*, 1951–1979. https://doi.org/10.1007/s00181-021-02071-x

Bouheni, F. Ben, Obeid, H., & Margarint, E. (2021). Nonperforming loan of European Islamic banks over the economic cycle. *Annals of Operations Research*, *313*, 773–808. https://doi.org/10.1007/s10479-021-04038-8

Butt, U., & Chamberlain, T. (2025). Performance of Islamic Banks During the COVID-19 Pandemic: An Empirical Analysis and Comparison with Conventional Banking. *Journal of Risk and Financial Management*. https://doi.org/10.3390/jrfm18060308

Carmona, P., Dwekat, A., & Mardawi, Z. (2022). No More Black Boxes! Explaining the Predictions of a Machine Learning XGBoost Classifier Algorithm in Business Failure. *Research in International Business and Finance*. https://doi.org/10.1016/j.ribaf.2022.101649

Černevičienė, J., & Kabašinskas, A. (2024). Explainable artificial intelligence (XAI) in finance: a systematic literature review. *Artificial Intelligence Review*, *57*. https://doi.org/10.1007/s10462-024-10854-8

Chen, H., Wang, T., & Yao, D. (2021). Financial Network and Systemic Risk—A Dynamic Model. *Production and Operations Management*, *30*, 2441–2466. https://doi.org/10.1111/poms.13384

De Lange, P., Melsom, B., Vennerød, C. B., & Westgaard, S. (2022). Explainable AI for Credit Assessment in Banks. *Journal of Risk and Financial Management*. https://doi.org/10.3390/jrfm15120556

Dwivedi, R., Dave, D., Naik, H., Singhal, S., Rana, O., Patel, P., Qian, B., Wen, Z., Shah, T., Morgan, G., & Ranjan, R. (2022). Explainable AI (XAI): Core Ideas, Techniques, and Solutions. *ACM Computing Surveys*, *55*, 1–33. https://doi.org/10.1145/3561048

Faizulayev, A. (2025). Empirical examination of ESG and fintech factors on financial sustainability: a comparative study of Islamic vs conventional banks in Islamic finance-oriented countries. *Asian Journal of Accounting Research*. https://doi.org/10.1108/ajar-09-2024-0382

Filippopoulou, C., Galariotis, E., & Spyrou, S. (2020). An early warning system for predicting systemic banking crises in the Eurozone: A logit regression approach. *Journal of Economic Behavior & Organization*. https://doi.org/10.1016/j.jebo.2019.12.023

Gafsi, N. (2025). Machine Learning Approaches to Credit Risk: Comparative Evidence from

Participation and Conventional Banks in the UK. *Journal of Risk and Financial Management*. https://doi.org/10.3390/jrfm18070345

Gao, R., Cui, S., Wang, Y., & Xu, W. (2025). Predicting financial distress in high-dimensional imbalanced datasets: a multi-heterogeneous self-paced ensemble learning framework. *Financial Innovation*, *11*. https://doi.org/10.1186/s40854-024-00745-w

Ghosh, S. (2025). A Novel Framework for Financial Cybersecurity and Fraud Detection Using XAI-RNN-SGRU. *IEEE Access*, *13*, 88134–88155. https://doi.org/10.1109/access.2025.3570216

Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., Scardapane, S., Spinelli, I., Mahmud, M., & Hussain, A. (2023). Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence. *Cognitive Computation*, *16*, 45–74. https://doi.org/10.1007/s12559-023-10179-8

Ijaz, S., Hassan, A., Tarazi, A., & Fraz, A. (2020). Linking Bank Competition, Financial Stability, And Economic Growth. *Journal of Business Economics and Management*. https://doi.org/10.3846/jbem.2020.11761

Kovalenko, Y., & Petruk, A. (2025). EARLY Response System In Ensuring Financial Stability Of Banks When Carrying Out Transactions With Derivative Financial Instruments. *Social Development: Economic and Legal Issues*. https://doi.org/10.70651/3083-6018/2025.5.04

Leelavathi, C., Malathi, K., Nithya, C., Gowd, A. B., & K.Gajendra. (2025). Federated Learning and Explainable AI: A Transparent and Secure Approach to Fraud Detection. *Power System Technology*. https://doi.org/10.52783/pst.1645

Liu, J., Li, C., Peng, O., Liu, J., & Wu, C. (2022). Interpreting the prediction results of the tree-based gradient boosting models for financial distress prediction with an explainable machine learning approach. *Journal of Forecasting*. https://doi.org/10.1002/for.2931

Liu, S. (2025). LTR-Net: A deep learning-based approach for financial data prediction and risk evaluation in enterprises. *PLOS One*, *20*. https://doi.org/10.1371/journal.pone.0328013

Majid, N. H. A., Yusof, R. M., Wahab, N. A., Rahim, M. A. A. A., Wiranatakusuma, D., & Tasman, S. M. (2025). Understanding resilience of Islamic banking amid COVID-19: a systematic literature review of banking, macroeconomic, and risk factors. *Journal of Islamic Accounting and Business Research*. https://doi.org/10.1108/jiabr-10-2023-0356

Martins, T., De Almeida, A. M., Cardoso, E., & Nunes, L. (2024). Explainable Artificial Intelligence (XAI): A Systematic Literature Review on Taxonomies and Applications in Finance. *IEEE Access*, *12*, 618–629. https://doi.org/10.1109/access.2023.3347028

Mehreen, M., Marimuthu, M., Karim, S. A. A., & Jan, A. (2020). Proposing a Multidimensional Bankruptcy Prediction Model: An Approach for Sustainable Islamic Banking. *Sustainability*. https://doi.org/10.3390/su12083226

Mohammad, S., Asutay, M., Dixon, R., & Platonova, E. (2020). Liquidity risk exposure and its determinants in the banking sector: A comparative analysis between Islamic, conventional and hybrid banks. *Journal of International Financial Markets, Institutions* *and Money*. https://doi.org/10.1016/j.intfin.2020.101196

Nallakaruppan, M., Chaturvedi, H., Grover, V., Balusamy, B., Jaraut, P., Bahadur, J., Meena, V., & Hameed, I. (2024). Credit Risk Assessment and Financial Decision Support Using Explainable Artificial Intelligence. *Risks*. https://doi.org/10.3390/risks12100164

Oino, I. (2021). Bank solvency: The role of credit and liquidity risks, regulatory capital and economic stability. *Banks and Bank Systems*. https://doi.org/10.21511/bbs.16(4).2021.08

Paltrinieri, A., Dreassi, A., Rossi, S., & Khan, A. (2020). Risk-adjusted profitability and stability of Islamic and conventional banks: Does revenue diversification matter? *Global Finance Journal*. https://doi.org/10.1016/j.gfj.2020.100517

Purnell, D., Etemadi, A., & Kamp, J. (2024). Developing an Early Warning System for Financial Networks: An Explainable Machine Learning Approach. *Entropy*, *26*. https://doi.org/10.3390/e26090796

Qian, H., Wang, B., Yuan, M., Gao, S., & Song, Y. (2021). Financial distress prediction using a corrected feature selection measure and gradient boosted decision tree. *Expert Syst. Appl.*, *190*, 116202. https://doi.org/10.1016/j.eswa.2021.116202

Saeed, W., & Omlin, C. (2021). Explainable AI (XAI): A Systematic Meta-Survey of Current Challenges and Future Opportunities. *ArXiv*, *abs/2111.06420*. https://doi.org/10.1016/j.knosys.2023.110273

Safdar, R., Fatima, A., & Sajid, M. (2024). Efficiency, liquidity risk and asset quality: a comparative analysis of Islamic and conventional banks in Pakistan. *Journal of Islamic Accounting and Business Research*. https://doi.org/10.1108/jiabr-03-2023-0079

Sun, B., & Liu, Y. (2025). A SEIQRS Model for Interbank Financial Risk Contagion and Rescue Strategies in Complex Networks. *Mathematics*. https://doi.org/10.3390/math13193059

Sun, T. (2025). Systemic risk of systemically important financial institutions in the post-2008 global financial crisis era: A tail risk network analysis. *Journal of Risk and Insurance*. https://doi.org/10.1111/jori.70023

Ullah, S., Ullah, A., & Zaman, M. (2024). Nexus of governance, macroeconomic conditions, and financial stability of banks: a comparison of developed and emerging countries. *Financial Innovation*, *10*. https://doi.org/10.1186/s40854-023-00542-x

Wang, T., Zhao, S., Zhu, G., & Zheng, H. (2021). A machine learning-based early warning system for systemic banking crises. *Applied Economics*, *53*, 2974–2992. https://doi.org/10.1080/00036846.2020.1870657

Weber, P., Carl, K., & Hinz, O. (2023). Applications of Explainable Artificial Intelligence in Finance—a systematic review of Finance, Information Systems, and Computer Science literature. *Management Review Quarterly*, *74*, 867–907. https://doi.org/10.1007/s11301-023-00320-0

Yin, S. (2025). Machine Learning Algorithms for Early Warning Systems: Predicting Systemic Financial Crises Through Non-Linear Econometric Models. *Computer Fraud and Security*.

https://doi.org/10.52710/cfs.613

Zhang, Z., Wu, C., Qu, S., & Chen, X. (2022). An explainable artificial intelligence approach for financial distress prediction. *Information Processing & Management*, *59*(4), 102988. https://doi.org/https://doi.org/10.1016/j.ipm.2022.102988