Research of Finance and Banking
p-ISSN/e-ISSN:2987-288X/2987-2871
Homepage: https://sanscientific.com/journal/index.php/rfb
4(1) 18-31 (2026)
https://doi.org/10.58777/rfb.v4i1.548

SAN Scientific

Research Article

Interpretable Machine Learning for Predicting Financial Distress in Emerging Market Insurance Sectors

Jahir Uddin Palas[1], Benazir Imam Majumder[2*]
[1,2] Dhaka University, Bangladesh

Received: 16-10-2025; Accepted: 03-05-2026

Abstract

This study develops an interpretable early-warning framework to predict financial distress among non-life insurers operating in a thin-premium market context. While prior studies widely rely on traditional models such as the Z-score, a critical research gap remains, as these models are not well-suited to the insurance industry due to its unique capital structures, regulatory requirements, and underwriting dynamics. Specifically, conventional distress prediction approaches tend to overlook operational characteristics such as reinsurance dependency, reserve adequacy, and expense management, which are central to insurer solvency. Addressing this gap, the study applies machine learning techniques combined with explainable artificial intelligence to enhance both predictive capability and transparency. Using firm-level panel data, the research incorporates key financial and operational indicators to construct a context-specific predictive framework. The methodology emphasizes balanced model evaluation, feature relevance, and interpretability to ensure practical applicability for supervisory authorities. By integrating explainability into predictive modeling, this study helps bridge the regulatory trust gap associated with black-box algorithms. The proposed framework offers a policy-relevant tool for early identification of vulnerable insurers and for facilitating timely intervention. Overall, this research advances the literature by aligning predictive accuracy with institutional usability in emerging insurance markets.

Keywords: financial distress prediction; non-life insurance; explainable machine learning; early-warning systems

JEL Classification: C53, G22, G33

How to cite: Palas, J. U., Majumder, B. I., (2026). Interpretable Machine Learning for Predicting Financial Distress in Emerging Market Insurance Sectors, *Research of Finance and Banking (RFB) 4(1)*, 18-31

Corresponding author: Benazir Imam Majumder (benazir.majumder@du.ac.bd)

1. Introduction

Financial distress prediction has become a critical issue in the modern insurance industry, especially in emerging markets where limited capital reserves and volatile underwriting conditions amplify systemic vulnerability. The collapse or distress of insurance companies can trigger severe economic and social consequences, including delayed claim settlements, erosion of consumer confidence, and fiscal burdens on governments (Ayinaddis & Tegegne, 2023; Kebede, Tesfaye, & Erana, 2024).

In many developing economies, insurance firms operate on thin premium bases and limited reinsurance coverage, leaving them more exposed to operational inefficiencies and market shocks. As the insurance sector plays a vital role in risk pooling and financial stability, the early detection of financial distress through predictive modeling is indispensable to ensure the industry's sustainability (Grize, Fischer, & Lützelschwab, 2020; Lokanan & Ramzan, 2024).

In recent years, technological advances and data availability have encouraged researchers to apply machine learning (ML) to financial distress prediction, improving accuracy and providing earlier warnings compared to traditional statistical approaches (Abrahamsen, Nylén-Forthun, Møller, de Lange, & Risstad, 2024; Dhamo, Gjeçi, Zibri, & Prendi, 2025). Traditional models such as Altman's Z-score and Ohlson's O-score, while foundational, rely heavily on linear relationships and are limited in their ability to capture complex interactions among financial ratios (Altman, 1968; Ohlson, 1980). In contrast, ensemble and boosting algorithms can capture nonlinear patterns, address data imbalance, and enhance sensitivity to rare events such as insurer distress without compromising interpretability when integrated with explainable AI tools (Bussmann et al., 2024; Kuizinienė, 2022). This advancement provides a strong rationale for applying ML-based frameworks in the insurance sector, particularly in emerging markets where data irregularities and limited historical distress events present unique challenges.

Despite the progress in financial distress modeling, most existing studies focus on banking, manufacturing, or publicly listed firms, with limited evidence from the insurance industry (Shetty, Musa, & Brédart, 2022). The insurance business exhibits distinctive operational structures, such as underwriting cycles, reserve management, and reinsurance dependency, which differentiate its distress dynamics from those of other industries (Bragoli et al., 2021; Valaskova, Kliestik, Svabova, & Adamko, 2018). Furthermore, studies targeting developing economies often overlook contextual factors such as data scarcity, weak governance, and market concentration that affect the robustness and applicability of prediction models (Ashraf & Vincent, 2021; Kebede et al., 2024). This gap underscores the need for context-specific research that balances predictive accuracy and interpretability, enabling regulators to make informed, data-driven decisions.

Another pressing issue is the limited interpretability of many ML-based distress models, which restricts their practical adoption by regulators and policymakers. While complex algorithms achieve superior statistical performance, they often operate as "black boxes," providing little insight into which financial indicators trigger distress warnings (El Madou, Marso, El Kharrim, & El Merouani, 2023; Hajek & Munk, 2023). Explainable Artificial Intelligence (XAI) frameworks, such as SHAP (Shapley Additive Explanations), offer a promising solution by translating algorithmic outputs into understandable financial indicators (Bussmann et al., 2024). This interpretability bridges the gap between academic innovation and regulatory usability, enabling supervisory agencies to act swiftly and confidently when early warning signals appear (Petropoulos, Siakoulis, Stavroulakis, & Vlachogiannakis, 2020).

In Bangladesh, the non-life insurance sector provides a particularly suitable environment to study financial distress due to its concentrated premium base, evolving regulatory structure, and exposure to macroeconomic instability. Many insurers operate on narrow margins and exhibit uneven underwriting performance, which makes the identification of distress-prone firms a regulatory priority (Abrahamsen et al., 2024; Kebede et al., 2024). The absence of prior empirical studies focused on Bangladesh's insurance sector further justifies this research. Examining how operational indicators such as management expense ratio, reinsurance intensity, and combined ratio contribute to early-warning detection will fill a crucial knowledge and policy gap (Ayinaddis & Tegegne, 2023; Grize et al., 2020).

This study, therefore, aims to develop an interpretable early-warning framework to predict financial distress among non-life insurers in Bangladesh using machine learning techniques. Specifically, it seeks to (1) evaluate the performance of several predictive algorithms logistic regression, K-nearest neighbors, support vector machines, random forest, and gradient-boosted trees in detecting distress;

(2) address class imbalance through oversampling methods; and (3) employ SHAP analysis to identify the most influential accounting indicators driving distress predictions (Lokanan & Ramzan, 2024; Dhamo et al., 2025). The combination of predictive performance and interpretability ensures that the results are both academically rigorous and operationally relevant for regulators.

Financial distress prediction in the Bangladeshi non-life insurance sector is a critical issue, not only because of market volatility but also due to its potential systemic implications for the broader economy. The failure of a single insurer can trigger cascading effects, including delayed claim settlements, erosion of public trust, disruption of risk-sharing mechanisms, and increased fiscal pressure on the government to stabilize affected policyholders. These risks are amplified in a thin-premium market characterized by limited capital buffers, concentrated market structures, and uneven underwriting practices, making the system highly sensitive to firm-level shocks. Despite the growing application of machine learning to financial distress prediction, most existing approaches operate as "black boxes," limiting their adoption by regulators who require transparent, explainable decision-making tools. This study addresses this dual challenge by proposing an interpretable machine learning framework that not only improves early-warning detection but also enhances regulatory confidence. By integrating SHAP-based explanations, the model translates complex algorithmic outputs into actionable financial indicators, thereby bridging the regulatory trust gap. This interpretability-driven approach represents the study's key novelty, as it aligns advanced predictive performance with the practical needs of supervisory authorities in emerging insurance markets.

2. Literature Review and Hypothesis

Literature Review

Financial Distress Theory

Financial distress theory explains a firm's inability to meet financial obligations without restructuring or external assistance. It arises from persistent operational inefficiency, declining profitability, and excessive leverage. While the foundational frameworks of Altman and Ohlson remain widely used, recent studies have re-evaluated and extended these models using updated datasets and modern analytical approaches (Altman et al., 2017; Liang et al., 2020). The classical Z-score model introduced discriminant analysis based on accounting ratios, whereas Ohlson’s probabilistic approach emphasized solvency, liquidity, and performance indicators. Contemporary research advances these frameworks by incorporating dynamic, multi-period, and machine learning–based perspectives, highlighting that financial distress evolves through both short-term liquidity pressures and long-term solvency deterioration (Shumway, 2001; Wang et al., 2025). In addition, recent literature stresses the importance of governance quality, operational efficiency, and market-specific conditions particularly in emerging economies in shaping distress outcomes. These developments reinforce the view that financial distress is not a single-point failure but a gradual and measurable process reflected in financial ratios and operational indicators over time.

Agency Theory

Agency theory suggests that conflicts between owners (principals) and managers (agents) can lead to inefficient resource allocation, excessive risk-taking, and moral hazard, which contribute to financial distress (Liang, Tsai, Lu, & Chang, 2020; Bussmann et al., 2024). In the insurance industry, such conflicts manifest as rising management expenses and suboptimal reinsurance policies when managerial incentives are not aligned with long-term solvency goals. Poor governance structures and ineffective oversight increase operational costs and weaken internal controls, thereby heightening distress risk (Ashraf & Félix, 2019; Balasubramanian et al., 2019). Hence, expense ratios and governance-related indicators often serve as proxies for agency inefficiencies in distress prediction models.

Signaling Theory

Signaling theory explains how firms communicate their financial health to external stakeholders through observable financial indicators. Healthy firms signal stability through profitability, adequate reserves, and consistent underwriting performance, while distressed firms may manipulate short-term signals such as reported capital adequacy to mask internal weaknesses (Abrahamsen, Nylén-Forthun, Møller, de Lange, & Risstad, 2024; Ashraf & Félix, 2019). In non-life insurance markets, manipulated signals often manifest as under-reserving or aggressive premium pricing, which temporarily boost solvency ratios but increase future claim liabilities (Petropoulos, Siakoulis, Stavroulakis, & Vlachogiannakis, 2020). Consequently, financial ratios that measure profitability, reserve adequacy, and underwriting efficiency are essential indicators of authentic versus deceptive signals in predicting distress.

Capital Structure and Contingent Claims Theory

Capital structure theory posits that a firm's mix of debt and equity affects its risk of distress due to differences in financial leverage and cost of capital. Highly leveraged firms face increased insolvency risk because fixed obligations magnify losses during downturns (Bragoli et al., 2021; Valaskova, Kliestik, Svabova, & Adamko, 2018). In insurance firms, capital adequacy reflects both solvency and regulatory compliance, while contingent claims theory interprets solvency as the outcome of option-like payoffs on assets and liabilities (Wang et al., 2025; Grize, Fischer, & Lützelschwab, 2020). These theories suggest that variables such as equity ratio and leverage serve as structural indicators of financial resilience or vulnerability to distress.

Underwriting Cycle Theory

Underwriting cycle theory explains periodic fluctuations in insurance profitability, driven by market competition, claim trends, and pricing strategies. During "soft" market phases, insurers reduce premiums to gain market share, increasing future claims exposure and reducing solvency margins (Grize et al., 2020; Bragoli et al., 2021). Firms with weak cost controls and aggressive growth strategies face an increased risk of financial distress when premium revenue fails to cover claims and expenses (Kebede, Tesfaye, & Erana, 2024; Ayinaddis & Tegegne, 2023). The combined ratio, reinsurance dependency, and management expense ratio thus become essential measures of underwriting efficiency and key predictors of early warning systems.

Resource-Based View (RBV) and Organizational Resilience Theory

The resource-based view (RBV) and organizational resilience theory emphasize the role of internal capabilities, governance, and operational efficiency in ensuring long-term financial stability (Valaskova et al., 2018; Liang et al., 2020). Firms that manage resources effectively such as human capital, underwriting expertise, and risk management systems are better equipped to withstand external shocks. In contrast, operational inefficiencies, weak governance, and poor expense management increase the risk of financial distress. These theories explain why managerial competence, cost control, and adaptive capacity are critical internal determinants of financial sustainability in the insurance industry (Kebede et al., 2024; Bussmann et al., 2024).

Hypothesis

Return on Assets (ROA) and Financial Distress.

Return on Assets (ROA) reflects a company's ability to generate profits from its total assets. Based on profitability and signaling theory, the higher the ROA, the better the company's financial condition because it indicates the efficiency of asset utilization in generating revenue (Ohlson, 1980; Abrahamsen et al., 2024). A decline in ROA is an early signal of financial distress, as insufficient profits fail to cover short-term liabilities. Previous research has shown that ROA is negatively related to financial distress in various industry contexts, including the insurance and banking sectors (Kebede et al., 2024; Ayinaddis & Tegegne, 2023; Wu, Yang, & colleagues, 2022; Shetty, Musa, & Brédart, 2022). Thus, a high level of profitability reduces the likelihood that a company will experience financial distress.

H1: Return on Assets (ROA) has a negative and significant effect on financial distress.

Equity Ratio to Financial Distress.

The equity ratio indicates the proportion of equity capital to total assets and serves as an indicator of solvency. According to capital structure theory, companies with a strong capital base are more resilient to economic shocks and the risk of loss (Altman, 1968; Bragoli et al., 2021). A low equity ratio indicates a high reliance on debt, thus increasing the potential for default. Previous research

confirms a negative relationship between the equity ratio and financial distress in the financial and insurance sectors. Petropoulos et al. (2020) found that capital adequacy is a key predictor of bank solvency, while Valaskova et al. (2018) and Kebede et al. (2024) showed that a low equity ratio accelerates the risk of bankruptcy. Furthermore, Liang et al. (2020) emphasized that good corporate governance can strengthen the protective effect of equity capital.
H2: The equity ratio has a negative and significant effect on financial distress.

Management Expense Ratio: Financial Distress.
According to agency theory, high management expenses reflect operational inefficiency and weak oversight of managerial behavior (Bussmann et al., 2024; Liang et al., 2020). In the insurance industry, a high management expense ratio indicates inefficient resource use and poor internal controls, which can worsen financial performance. Empirical studies support the important role of this ratio in detecting financial distress. Kebede et al. (2024) found that the management expense ratio was a key predictor of financial distress in Ethiopian insurance companies. Similar findings were also demonstrated by Grize, Fischer, and Lützelschwab (2020), who found that unmanageable management expenses were an early signal of a company's inability to pay claims. Furthermore, Ayinaddis and Tegegne (2023) and Lokanan and Ramzan (2024) confirmed that operational efficiency is a key indicator of financial health in the insurance sector.
H3: The Management Expense Ratio has a positive and significant effect on financial distress.

Reinsurance Ratio to Financial Distress.
The reinsurance ratio measures a company's dependence on reinsurance to transfer underwriting risk. According to contingent claims theory and the underwriting cycle theory, excessive reliance on reinsurance can indicate internal capital weakness and reduce potential net income (Grize et al., 2020; Bragoli et al., 2021). Empirically, Kebede et al. (2024) found that high reliance on reinsurance increases the likelihood of financial distress by reducing premium retention. Ayinaddis and Tegegne (2023) added that companies with a high proportion of reinsurance are more vulnerable to external shocks and the risk of default by reinsurance counterparties. Furthermore, research by Shetty et al. (2022) and Petropoulos et al. (2020) also shows that excessive reinsurance strategies weaken a company's financial stability.
H4: The reinsurance ratio has a positive and significant effect on financial distress.

Lagged Combined Ratio on Financial Distress.
The lagged combined ratio reflects underwriting efficiency in the previous period. Based on underwriting cycle theory, poor underwriting performance has a lasting effect on solvency because past losses can reduce a company's ability to cover future claims (Ayinaddis & Tegegne, 2023; Grize et al., 2020). Previous research indicates that this ratio is a leading indicator of financial distress in the insurance sector. Kebede et al. (2024) found that a high combined ratio reflects underwriting losses and is an early signal of distress. Wu et al. (2022) and Abrahamsen et al. (2024) also confirmed that historical underwriting performance influences long-term financial stability. Therefore, this ratio is used as a primary predictor in early warning models.
H5: The lagged combined ratio has a positive and significant effect on financial distress.

Conceptual Framework

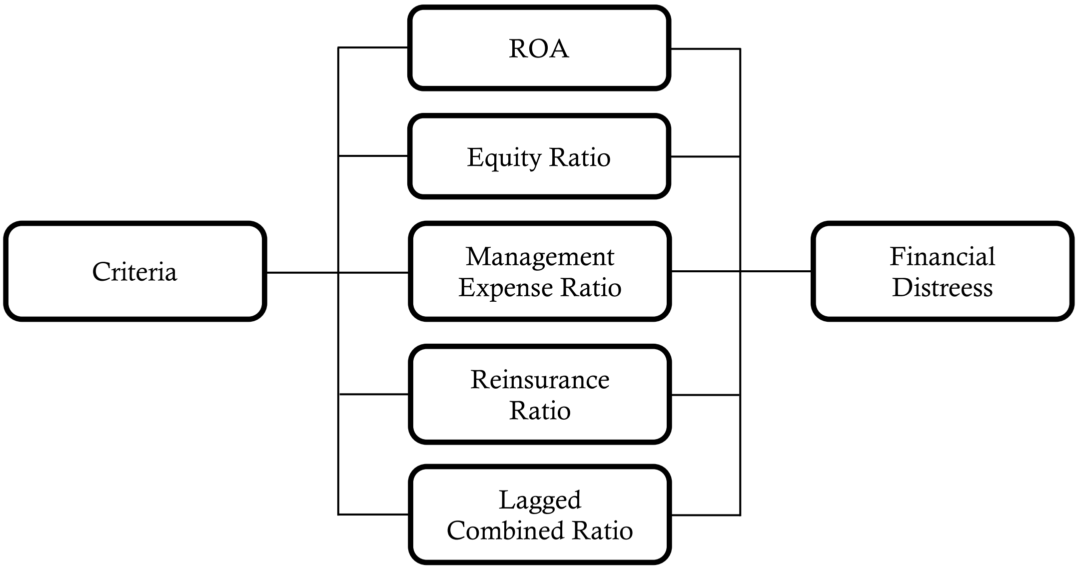

Figure 1. Conceptual Framework

The conceptual framework illustrates the relationship between five independent variables Return on Assets (ROA), Equity Ratio, Management Expense Ratio, Reinsurance Ratio, and Lagged Combined Ratio and the dependent variable, Financial Distress. Each independent variable has an arrow pointing toward Financial Distress, indicating a direct influence. This model shows that profitability and capital strength can reduce the risk of financial distress, while management inefficiency, reinsurance dependence, and poor underwriting performance increase the likelihood of financial distress in non-life insurance companies.

3. Data and Method

We developed a predictive framework to classify financial distress among general insurance companies in Bangladesh using firm-level accounting ratios and supervised machine learning. This section outlines our labeling strategy, feature selection, model set, evaluation protocol, and robustness checks. We designed the workflow to reflect regulatory requirements, emphasizing out-of-time validation, interpretable predictors, and sensitivity analyses that inform operational implementation.

This dataset covers all general insurance companies operating in Bangladesh from 2014 to 2024. Financial data were collected from each insurance company's annual report and the Bangladesh Insurance Association (BIA) yearbook. The final dataset consists of 506 company-year observations across 46 companies, including 45 private insurers and one public insurer. However, data for three companies were unavailable for 2024, as their annual reports had not been published as of 31 August 2025. The number of missing values was less than 1% of the total observations.

Data Splitting, Preprocessing, and Balancing

We divided the panel into two sets: a training set (2014-2022) and a test set (2023-2024). We used the training set for learning from observations and the testing set for out-of-sample prediction simulations. Due to the inherent class imbalance caused by the small number of catastrophe insurance companies, we applied the Synthetic Minority Oversampling Technique (SMOTE) to the training dataset to balance the classes and improve the model's ability to learn from limited catastrophe observations without introducing data leakage bias. Furthermore, we standardized all features using StandardScaler from the scikit-learn library in Python to ensure fair and consistent model training. Using this normalization process, we set all features to zero and their standard deviations to 1. This step is crucial for models sensitive to feature scale.

Furthermore, we excluded the current-year combined ratio from the feature set to prevent target leakage. Instead, we used the one-period lagged combined ratio as a predictor of financial distress in the main models. Therefore, this adjustment reduces the total number of observations to 460, covering the period 2015 to 2024.

Hyperparameter Tuning and Cross-Validation

We optimized hyperparameters using a 5-fold cross-validation grid search on the training data to improve model performance. This approach divides the training data into five folds, trains the model on four folds, and evaluates the model's performance on the remaining fold. This process is repeated across all combinations within the specified hyperparameter grid to identify the best hyperparameters based on the F1 score. We then retrained each model with optimal hyperparameters on the full training set using a similar processing pipeline, including mean imputation, feature scaling, and SMOTE.

Robustness Checks

To assess the reliability of the best predictive models, we applied several robustness checks. First, we used alternative re-sampling methods on the best-performing models. Here, we tested Adaptive Synthetic Sampling (ADASYN) and no re-sampling, in addition to SMOTE, to assess the impact of different class imbalance handling techniques. Second, we performed feature reduction based on SHAP values. We retrain the model on the reduced feature set to examine its results after excluding low-importance and highly correlated variables. Third, we maintained consistent time-based validation. We compared performance across these variations using standard metrics, with the results discussed in the results section and visualizations provided in the appendix.

4. Results

Summary Statistics and Preliminary Analyses

The dataset comprises 506 company-year observations from 46 non-life insurers in Bangladesh over the period 2014 to 2024. Among 506 observations, three of the company's 2024 data were imputed with mean values due to missing values. Using the criterion of a combined ratio greater than 1, Table 6 shows that 404 observations (79.84%) correspond to non-distressed insurers, while 102 (20.16%) correspond to distressed insurers. Moreover, the yearly distribution of distress shows notable year-to-year fluctuations in the proportion of the distressed insurers in Bangladesh from 2014 to 2024. Table 1 presents the percentage of distress company-year observations, along with their actual counts, for the combined ratio criteria.

Table 1. Yearly Distribution of Financial Distress

Year	Non-distressed	Distressed	Distressed (%)
2014	31	15	32.61
2015	34	12	26.09
2016	38	8	17.39
2017	41	5	10.87
2018	34	12	26.09
2019	36	10	21.74
2020	37	9	19.57
2021	43	3	6.52
2022	36	10	21.74
2023	36	10	21.74
2024	38	8	17.39
Total	**404**	**102**	**100%**

Source: Processed data (2025)

Yearly distress patterns indicate that 2014 recorded the highest proportion of distressed insurers, while 2021 showed the lowest, highlighting the influence of macroeconomic conditions, regulatory changes, and firm-level strategies. This variation supports the use of time-variant factors, such as the lagged combined ratio, in predicting distress. Table 2 shows that distressed insurers generally have lower profitability, weaker capital positions, and slower growth, alongside higher management expenses, leverage, reinsurance dependence, and lagged combined ratios. The results further indicate that profitability, operational costs, and reliance on reinsurance are key factors associated with financial distress.

Table 2. Comparative Analysis of Group Means

Feature	Mean (Distressed)	Mean (non-distressed)	Mean Difference	P-value
ROA	0.0585	0.0708	-0.0123	**0.0046***
Equity Ratio	0.3294	0.3335	-0.0042	0.8528
Reinsurance Ratio	0.4645	0.4010	0.0635	**0.0000***
Mgt. Exp Ratio	0.6203	0.4192	0.2011	**0.0000***
Investment Yield	0.499	0.4461	0.0528	0.3346
URR Ratio	0.4181	0.4055	0.0126	0.2105
Company Size	3.1922	3.1961	-0.0038	0.9119
Leverage	0.6706	0.6665	0.0042	0.8528
Premium Growth	0.0626	0.1299	-0.0674	0.1136
Lagged Combined Ratio	0.8198	0.7843	0.0355	0.4581

Source: Processed data (2025)

We conduct a Variance Inflation Factor (VIF) analysis to assess multicollinearity among the features selected for predicting financial distress. Table 3 shows that two important features, Leverage and Equity ratio, exhibit high VIF values. However, we retain these variables because our goal is to achieve predictive accuracy in modeling financial distress rather than to interpret coefficients. Additionally, we use logistic regression as a benchmark model, and the three-based models (Random Forest and XGBoost) are robust to multicollinearity by design. As these models form the core of our analysis, we expect that multicollinearity will not adversely affect their predictive performance.

Table 3. Multicollinearity

Feature	VIF
ROA	1.109765
Equity Ratio	15.72185
Reinsurance Ratio	1.172303
Mgt Exp Ratio	1.236439
Investment Yield	1.090808
URR Ratio	1.14585
Company Size	1.644624
Leverage	89.15425
Premium Growth	1.015793
Lagged Combined Ratio	1.020458

Source: Processed data 2025

In summary, these descriptive statistics and preliminary analyses provide a solid foundation for subsequent predictive modeling with benchmark, classical, and ensemble models.

Model Performance Using Default Settings

Following data preprocessing, five supervised machine learning models logistic regression, KNN, SVM, random forest, and XGBoost are applied using default hyperparameters. Missing values are imputed using the mean, and SMOTE is used to address class imbalance, as distressed firms constitute a minority. The models are trained on 2014–2022 data and tested on 2023–2024 data. Table 4 presents the performance results, while detailed classification outcomes are provided in the confusion matrices in the Online Appendix.

Table 4. Default Model Performance Summary

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Logistic Regression	0.7065	0.3714	0.7222	0.4906	0.7538
K-Nearest Neighbors	0.7609	0.4231	0.6111	0.5000	0.7395
Support Vector Machine	0.7935	0.4815	0.7222	0.5778	0.8071
Random Forest	0.8152	0.5238	0.6111	0.5641	0.8431
XGBoost	0.7717	0.4545	0.8333	0.5882	0.8468

Source: Processed data (2025)

Table 4 shows that logistic regression, as a benchmark model, performs moderately in detecting financial distress, with reasonable accuracy and recall but relatively low precision, suggesting some misclassification of healthy firms. Among non-ensemble methods, SVM provides more balanced performance than KNN, with higher accuracy and stronger sensitivity to distressed cases, while KNN shows comparatively lower recall. For ensemble models, Random Forest delivers robust, stable performance, but XGBoost outperforms all models, achieving the highest recall and strong overall predictive performance. Overall, all models exhibit acceptable discrimination (ROC-AUC > 0.70), though precision remains lower than recall due to class imbalance. Based on F1-score and ROC-AUC, XGBoost performs best under default settings.

Model Tuning and Evaluation

We apply 5-fold cross-validation and grid search for hyperparameter tuning as part of the robustness check. Grid search tests various parameter combinations, while cross-validation ensures reliable evaluation by iteratively training and validating the model across data splits. The optimal parameters are selected based on the average F1-score. Table 5 presents the best hyperparameters and their corresponding performance.

Table 5 Model Performance and Best Hyperparameters

Model	Best Hyper-parameters	CV F1 Score	Default Model F1 Score	Tuned Model F1 Score
Logistic Regression	Regularization strength (C) = 0.01, Penalty type (L1/L2) = L2, Solver choice = *lbfgs*	0.5095	0.4906	0.4815
K-Nearest Neighbors	Number of neighbors = 9, Distance metric = Manhattan, Weight function = Distance	0.5601	0.5000	0.5833
Support Vector Machine	Regularization parameter (C) = 1, Gamma = Scale, Kernel type = RBF	0.5215	0.5778	0.5000
Random Forest	Number of trees = 200, Maximum tree depth = None, Minimum samples per split = 2	0.5724	0.5641	0.5714
XGBoost	Learning rate = 0.01, Maximum tree depth = 3, Number of Estimator = 50, Subsample ratio =1	0.6300	0.5882	0.5614

Source: Processed data (2025)

Table 6 shows mixed results after hyperparameter tuning. Logistic regression remains relatively unchanged, indicating limited improvement. The tuned KNN model demonstrates notable gains in accuracy and recall, reflecting improved predictive performance. In contrast, the tuned SVM slightly underperforms its default version, achieving similar recall but lower overall accuracy.

Table 6. Tuned Model Performance Summary

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Logistic Regression	0.6957	0.3611	0.7222	0.4815	0.7515
K-Nearest Neighbors	0.7826	0.4667	0.7778	0.5833	0.8078
Support Vector Machine	0.7174	0.3824	0.7222	0.5000	0.7538
Random Forest	0.8043	0.5000	0.6667	0.5714	0.8296
XGBoost	0.7283	0.4103	0.8889	0.5614	0.8641

Source: Processed data (2025)

Hyperparameter tuning improves the performance of tree-based models. Random Forest shows higher recall and F1-score despite a slight decline in ROC-AUC, while XGBoost maintains strong performance with the highest recall and ROC-AUC, confirming its effectiveness. The confusion matrices further indicate improved identification of distressed firms after tuning. Overall, tuning has varying effects across models, generally increasing recall and precision at a minor cost to accuracy. Ensemble models remain robust, with XGBoost consistently outperforming others. The ROC curves also highlight improved discrimination, underscoring the importance of hyperparameter optimization to enhance prediction reliability.

Feature Importance in Predictive Analysis

The XGBoost model demonstrates consistent performance across both default and tuned settings. To identify key drivers of financial distress, feature importance is assessed using SHAP, thereby enhancing model interpretability. The results indicate that management expense ratio, reinsurance ratio, and lagged combined ratio are the most influential predictors, as presented in Table 7 and the feature importance plot in the Online Appendix.

Table 7. Shapley Additive Explanations

Feature	Mean SHAP Value	Importance %
Management Expense Ratio	0.89	43.53
Reinsurance Ratio	0.32	15.63
Lagged Combined Ratio	0.25	12.25
ROA	0.20	9.82
URR Ratio	0.12	6.03
Company Size	0.10	4.84
Premium Growth	0.08	3.78
Investment Yield	0.07	3.45
Equity Ratio	0.01	0.65
Leverage	0.00	0.02

Source: Processed data (2025)

The SHAP results show that the management expense ratio is the most dominant predictor of financial distress, contributing the largest share to the model's decisions, indicating that higher operating costs increase distress risk. The reinsurance ratio and the lagged combined ratio also play significant roles, underscoring the importance of risk transfer practices and past underwriting performance. Together, these three variables account for the majority of the model's predictive power. While profitability and reserve adequacy contribute moderately, variables such as company size, premium growth, and investment yield have limited influence, and equity ratio and leverage show minimal impact. Overall, the findings emphasize that expense efficiency, underwriting discipline, and reinsurance strategy are key early warning indicators of financial distress.

Robustness Check

We conduct several robust checks to ensure the reliability and stability of the XGBoost model. The reliability test evaluates model performance under varying class-imbalance handling strategies using a reduced feature set, eliminating financial indicators with lower mean SHAP values. Table 8 shows the result of the robustness check.

Table 8. Robustness Check

Re-sampling	Feature Set	Accuracy	Precision	Recall	F1-Score	ROC-AUC
SMOTE	Full Features	0.793	0.485	0.889	0.627	0.865
ADASYN	Full Features	0.783	0.471	0.889	0.615	0.865
None	Full Features	0.837	0.571	0.667	0.615	0.856
SMOTE	Reduced Features	0.761	0.444	0.889	0.593	0.884
ADASYN	Reduced Features	0.750	0.429	0.833	0.566	0.860
None	Reduced Features	0.837	0.579	0.611	0.595	0.843

Source: Processed data (2025)

We compare model performance across three re-sampling strategies: SMOTE, ADASYN, and no re-sampling. SMOTE and ADASYN improve recall but reduce precision, whereas no re-sampling yields higher accuracy and precision at the expense of lower recall, highlighting the trade-off in detecting minority distress cases. These results support the use of SMOTE for better identification of distressed firms. Excluding low-importance variables for feature reduction results in minimal performance loss, confirming their limited predictive value. Overall, the robustness checks demonstrate that the XGBoost model remains stable and reliable across different preprocessing scenarios, supporting its use for early warning systems.

5. Discussion

Effect of Return on Assets (ROA) on Financial Distress

The finding that Returns on Assets (ROA) negatively affect financial distress indicates that more profitable firms are less likely to experience financial instability. This aligns with profitability theory, which suggests that efficient asset utilization enhances earnings and liquidity, enabling firms to meet financial obligations successfully. The result reinforces the perspectives of Ohlson (1980) and Altman (1968) that profitability serves as a buffer against financial shocks. Similar findings by Kebede, Tesfaye, and Erana (2024) and Ayinaddis and Tegegne (2023) showed that higher ROA significantly reduces the likelihood of financial distress in insurance firms. In the context of non-life insurers in Bangladesh, this relationship underscores the importance of operational efficiency and effective investment management in maintaining solvency. Profitability thus serves as an early indicator of resilience, underscoring that sustained earnings power is vital for long-term financial health.

Effect of Equity Ratio on Financial Distress

The negative relationship between equity ratio and financial distress suggests that stronger capitalization enhances the financial stability of non-life insurers. According to capital structure theory, firms with greater equity financing have greater shock-absorption capacity and lower debt-service pressure (Bragoli et al., 2021). The results are consistent with previous studies by Petropoulos, Siakoulis, Stavroulakis, and Vlachogiannakis (2020) and Valaskova, Kliestik, Svabova, and Adamko (2018), which found that adequate capital reduces insolvency risk in financial institutions. This implies that the equity ratio is a critical indicator of risk-bearing capacity and regulatory solvency. In developing markets such as Bangladesh, where financial systems are often undercapitalized, maintaining a robust equity base is essential to mitigate exposure to market volatility and unexpected claims. The findings confirm that a firm's capital adequacy plays a central role in reducing the probability of distress.

Effect of Management Expense Ratio on Financial Distress

The positive and significant effect of the management expense ratio on financial distress demonstrates that operational inefficiency and excessive administrative spending increase the risk of financial instability. From an agency theory perspective, this result indicates misalignment between managerial actions and shareholder interests, as inefficient cost management erodes profitability and solvency (Liang, Tsai, Lu, & Chang, 2020). Empirical evidence from Kebede et al. (2024) and Grize, Fischer, and Lützelschwab (2020) supports this finding, emphasizing that higher management expenses reflect weak internal controls and poor governance practices. In the context of non-life insurance, excessive management costs may also signal ineffective underwriting strategies or misallocation of resources. Thus, firms that fail to optimize their operational expenditure are more vulnerable to liquidity shortages and long-term financial distress. The results underline the importance of efficient managerial control systems in ensuring organizational sustainability.

Effect of Reinsurance Ratio on Financial Distress
The results showing a positive relationship between the reinsurance ratio and financial distress indicate that over-reliance on reinsurance can undermine a firm's financial stability. Although reinsurance is intended to transfer risk and stabilize earnings, excessive reliance can reduce retained premiums and profitability (Grize et al., 2020). This finding aligns with Ayinaddis and Tegegne (2023) and Kebede et al. (2024), who reported that insurance firms with high reinsurance ratios are more prone to financial distress due to reduced revenue retention and exposure to counterparty risk. Theoretically, contingent claims and underwriting cycle theories suggest that over-reinsurance may signal weak internal risk management and insufficient capital to absorb shocks. In emerging markets, such patterns often reflect limited underwriting capacity rather than prudent risk sharing. Therefore, while reinsurance remains an essential risk mitigation tool, its overuse may paradoxically amplify financial vulnerability if not balanced with sufficient retained earnings.

Effect of Lagged Combined Ratio on Financial Distress
The positive relationship between the lagged combined ratio and financial distress indicates that past underwriting performance significantly affects a firm's current solvency. Underwriting cycle theory explains that losses from prior periods create a cumulative financial burden that weakens liquidity and profitability over time (Grize et al., 2020). Consistent with Kebede et al. (2024) and Wu, Yang, and colleagues (2022), this study finds that firms with consistently high combined ratios face higher distress risks due to persistent inefficiencies in risk assessment and claims management. In the non-life insurance sector, underwriting discipline directly determines long-term financial health. The persistence of high combined ratios suggests that firms may be engaging in aggressive premium competition or underpricing, which eventually erodes financial reserves. The finding emphasizes the need for prudent underwriting strategies and accurate risk pricing to sustain profitability and avoid cumulative financial distress.

6. Conclusion
This study demonstrates that gradient-boosted tree models, when integrated with SHAP-based interpretability, provide a robust and practical alternative to traditional logistic regression for predicting financial distress in emerging insurance markets. Beyond improving predictive performance, the framework enhances transparency by translating complex algorithmic outputs into clear and actionable indicators, thereby addressing the regulatory trust gap associated with black-box models. From a theoretical perspective, the findings validate the increasing relevance of operational metrics such as expense efficiency, underwriting performance, and reinsurance dependence over conventional financial ratios in explaining distress dynamics within thin-premium environments. This shifts the analytical paradigm from a "solvency-only" approach to a more comprehensive "efficiency-driven" monitoring framework, offering a deeper understanding of how financial vulnerability develops gradually rather than as a sudden event.

Economically and from a policy standpoint, the study provides a "policy-ready pipeline" that can support early intervention by regulatory authorities such as the Insurance Development and Regulatory Authority (IDRA) in Bangladesh. The results indicate that financial distress in the insurance sector is a cumulative process driven by persistent operational inefficiencies and weak underwriting discipline, which, if left unaddressed, may propagate systemic risk across the financial system. By delivering timely, interpretable early-warning signals, the proposed model enables regulators to detect vulnerable insurers more effectively, allocate supervisory resources efficiently, and implement preventive measures to mitigate potential contagion.

Recommendation

Regulators are advised to implement a dynamic threshold for monitoring management expense ratios, flagging firms that exceed a specified margin above the industry median for early supervisory review or a "soft audit." This approach enables proactive oversight without imposing excessive regulatory burden. In parallel, insurers should prioritize digital transformation and operational efficiency initiatives to reduce management costs and improve underwriting discipline. Strengthening internal controls, optimizing resource allocation, and adopting data-driven decision systems can significantly mitigate early signs of financial distress. Together, these measures support a more resilient insurance sector by aligning regulatory vigilance with firm-level efficiency improvements.

Limitations and avenues for future research

This study has several limitations that open avenues for future research. First, the analysis is confined to a single country context, which may limit the generalizability of the findings across different regulatory and market environments. Second, the model relies primarily on accounting-based indicators, potentially overlooking macroeconomic and behavioral factors that may influence financial distress. Future research should extend this framework to other South Asian markets, such as Pakistan and Sri Lanka, to assess its geographical robustness and adaptability. Additionally, incorporating cross-country datasets, stress-testing scenarios, and alternative data sources could further enhance predictive performance and policy relevance.

References

Abrahamsen, N.-G. B., Nylén-Forthun, E., Møller, M., de Lange, P. E., &Risstad, M. (2024). Financial distress prediction in the Nordics: Early warnings from machine learning models. *Journal of Risk and Financial Management, 17*(10), 432. https://doi.org/10.3390/jrfm17100432

Altman, E. I. (1968). Financial ratios, discrimination analysis and the prediction of corporate bankruptcy. *The Journal of Finance, 23*(4), 589–609.

Altman, E. I., Iwanicz-Drozdowska, M., Laitinen, E. K., & Suvas, A. (2017). Financial distress prediction in an international context: A review and empirical analysis of Altman's Z-score model. *Journal of International Financial Management & Accounting, 28*(2), 131–171. https://doi.org/10.1111/jifm.12053

Ashraf, S. H., & Vincent, T. N. (2021). The role of board independence and ownership structure in improving the efficacy of corporate financial distress prediction model: Evidence from India. *Journal of Risk and Financial Management, 14*(7), 333.

Ayinaddis, S. G., & Tegegne, H. G. (2023). Uncovering financial distress conditions and their determining factors on insurance companies in Ethiopia. *PLOS ONE, 18*(10), e0292973. https://doi.org/10.1371/journal.pone.0292973

Balasubramanian, S. A., & colleagues. (2019). Modeling corporate financial distress using financial and non-financial variables: The case of Indian listed companies. *International Journal of Law and Management, 61*(3–4), 457–477. https://doi.org/10.1108/IJLMA-04-2018-0078

Bragoli, D., Ferretti, C., Ganugi, P., Marseguerra, G., & colleagues. (2021). Machine-learning models for bankruptcy prediction: Do industrial variables matter? *Journal of Financial Studies, 17*(2). https://doi.org/10.1080/17421772.2021.1977377

Bussmann, N., et al. (2024). Explainable machine learning to predict the cost of capital. *Frontiers in Artificial Intelligence*. https://doi.org/10.3389/frai.2024.1466321

Dhamo, Z., Gjeçi, A., Zibri, A., & Prendi, X. (2025). Business distress prediction in Albania: An analysis of classification methods. *Journal of Risk and Financial Management, 18*(3), 118. https://doi.org/10.3390/jrfm18030118

El Madou, K., Marso, S., El Kharrim, M., & El Merouani, M. (2023). Evolutions in machine learning technology for financial distress prediction: A comprehensive review and comparative analysis. *Expert Systems, 40*(5). https://doi.org/10.1111/exsy.13485

Grize, Y.-L., Fischer, W., &Lützelschwab, C. (2020). Machine learning applications in non-life insurance. *Applied Stochastic Models in Business and Industry, 36*(5), 523–537. https://doi.org/10.1002/asmb.2543

Hajek, P., & Munk, M. (2023). Speech emotion recognition and text sentiment analysis for financial distress prediction. *Neural Computing and Applications, 35*, 21463–21477. https://doi.org/10.1007/s00521-023-08470-8

Hanafy, M., & Ming, R. (2021). Machine learning approaches for big data in auto insurance. *Risks, 9*(2), 42. https://doi.org/10.3390/risks9020042

Kebede, T. N., Tesfaye, G. D., & Erana, O. T. (2024). Determinants of financial distress: Evidence from insurance companies in Ethiopia. *Journal of Innovation and Entrepreneurship, 13*, Article 17. https://doi.org/10.1186/s13731-024-00369-5

Kuizinienė, D. (2022). Systematic review of financial distress identification using artificial intelligence methods. *International Journal of Financial Studies, 10*(4). https://doi.org/10.1080/08839514.2022.2138124

Lessmann, S., Baesens, B., Seow, H.V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update. *European Journal of Operational Research, 247*(1), 124–136. https://doi.org/10.1016/j.ejor.2015.05.030

Liang, D., Tsai, C.F., Lu, H.Y. R., & Chang, L. S. (2020). Combining corporate governance indicators with stacking ensembles for financial distress prediction. *Journal of Business Research, 120*, 137–146. https://doi.org/10.1016/j.jbusres.2020.07.052

Lokanan, M. E., & Ramzan, S. (2024). Predicting financial distress in TSX-listed firms using machine learning algorithms. *Frontiers in Artificial Intelligence, 7*, Article 1466321. https://doi.org/10.3389/frai.2024.1466321

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research, 18*(1), 109–131.

Petropoulos, A., Siakoulis, V., Stavroulakis, E., &Vlachogiannakis, N. E. (2020). Predicting bank insolvencies using machine learning techniques. *International Journal of Forecasting, 36*(3), 1092–1113. https://doi.org/10.1016/j.ijforecast.2019.11.005

Samitas, A., Kampouris, E., &Kenourgios, D. (2020). Machine learning as an early warning system for predicting financial crises. *International Review of Financial Analysis, 71*, 101507. https://doi.org/10.1016/j.irfa.2020.101507

Shetty, S., Musa, M., &Brédart, X. (2022). Bankruptcy prediction using machine learning techniques. *Journal of Risk and Financial Management, 15*(1), 35. https://doi.org/10.3390/jrfm15010035

Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model. *The Journal of Business, 74*(1), 101–124.

Tang, X., Li, S., Tan, M., & Shi, W. (2020). Incorporating textual and management factors into financial distress prediction: A comparative study of machine learning methods. *Journal of Forecasting, 39*(5), 769–787. https://doi.org/10.1002/for.2661

Valaskova, K., Kliestik, T., Svabova, L., & Adamko, P. (2018). Financial risk measurement and predictive modeling for the sustainable development of business entities using regression analysis. *Sustainability, 10*(7), 2144. https://doi.org/10.3390/su10072144

Wang, C., Gong, P., Li, J., & Wang, Z. (2025). Corporate financial distress prediction with multiperiod annual report data: A fusion deep neural network model. *PLOS ONE, 20*(9). https://doi.org/10.1371/journal.pone.0333064

Wu, D., Yang, H., & colleagues. (2022). Financial distress prediction using an integrated Z-score and a multilayer perceptron neural network. *Decision Support Systems, 159*, 113814. https://doi.org/10.1016/j.dss.2022.113814

Zhao, Q., Xu, W., & colleagues. (2023). Predicting financial distress of Chinese listed companies using machine learning: To what extent does textual disclosure matter? *International Review of Financial Analysis, 89*, 102770. https://doi.org/10.1016/j.irfa.2023.102770

Zizi, Y., Jamali-Alaoui, A., El Goumi, B., Oudgou, M., & El Moudden, A. (2021). An optimal model of financial distress prediction: A comparative study between neural networks and logistic regression. *Risks, 9*(11), 200. https://doi.org/10.3390/risks9110200