Proceedings of the 4th International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/2026.35021

An Explainable Machine Learning Framework for Early Warning of Profitability Downside Risk in Listed Companies

Xiaoqian Lin

School of Accounting, Guangdong University of Finance & Economics, Guangzhou, China
linxiaoqian586@foxmail.com

Abstract. Profitability deterioration often appears before bankruptcy, default, or formal financial distress. This paper develops an explainable machine learning framework to identify listed companies with high next-year profitability downside risk. Using a public firm-year panel constructed from SEC Financial Statement Data Sets and Stooq historical stock price data, this study examines U.S. non-financial listed companies from 2015 to 2024. High-risk observations are defined as firms whose next-year ROA change falls in the bottom 30 percent within the same industry-year group. Logistic Regression, Random Forest, XGBoost, and LightGBM are evaluated under a chronological validation design. XGBoost performs best in the out-of-sample test set, with an AUC of 0.836 and a PR-AUC of 0.653. SHAP results indicate that revenue growth, operating margin, leverage, ROA, operating cash flow growth, and stock volatility are the main risk drivers.

Keywords: Explainable Machine Learning, Profitability Downside Risk, Listed Companies, Financial Early Warning, SHAP, XGBoost, Decision Support

1. Introduction

Financial early-warning research has traditionally focused on bankruptcy, default, and financial distress, but these outcomes often appear after firms have already experienced operating deterioration. For managers and external stakeholders, earlier signals are more valuable. This paper therefore shifts the prediction object to profitability downside risk, defined as a significant decline in next-year ROA relative to firms in the same industry-year group. This setting keeps the warning signal close to operating performance while controlling for industry cycles.

Machine learning is suitable for this task because profitability changes may result from nonlinear interactions among growth, margins, liquidity, leverage, efficiency, and market information. Since financial early-warning systems also require interpretability, this study compares common machine learning models under a chronological split and uses SHAP to link model predictions with economic meanings.

2. Literature review

Early financial warning models originated from accounting-ratio approaches. Beaver showed that financial ratios predict future failure [1], Altman developed a multivariate discriminant model for

DOI: 10.54254/2755-2721/2026.35021

bankruptcy prediction [2], and Ohlson introduced probabilistic modeling to this field [3]. Later studies further incorporated market and accounting signals into distress-risk analysis [4]. These works confirm the importance of profitability, leverage, liquidity, and firm size, but they mainly focus on severe financial failure.

Machine learning has become increasingly useful for financial prediction because ensemble and boosting methods can capture nonlinear patterns in tabular data. Random Forest reduces variance through bagging and feature randomization [5], while gradient boosting builds additive trees to correct previous errors [6]. XGBoost improves scalability and regularization [7], and LightGBM supports efficient learning on structured datasets [8]. These methods are suitable for listed-company risk screening, where financial indicators often interact in complex ways.

Interpretability is also essential for financial decision-making. Even accurate models may be difficult to adopt if users cannot understand why a firm is flagged. SHAP, based on Shapley values [9], provides a unified additive feature-attribution framework [10]. This paper uses SHAP to link model outputs with economic meanings such as demand weakening, cost pressure, leverage burden, and cash flow deterioration.

3. Data and variable construction

The study uses a public dataset constructed from SEC Financial Statement Data Sets and Stooq historical stock price data for U.S. non-financial listed companies. The raw data cover 2015 to 2025, while labeled firm-year observations are constructed for 2015 to 2024. Financial firms are excluded due to their distinct balance-sheet structures and regulatory frameworks. Missing core accounting variables are removed, and continuous variables are winsorized at the 1st and 99th percentiles.

The final dataset contains 6,240 firm-year observations from U.S. non-financial companies, classified by two-digit SIC industries. To mimic forecasting, observations from 2015 to 2021 are used for training, 2022 for validation, and 2023 to 2024 for testing, which helps reduce look-ahead bias.

Profitability downside risk is defined by next-year ROA change, where ROA equals net income divided by total assets. A firm-year observation is labeled as high risk if its next-year ROA change falls into the lowest 30 percent within the same industry-year group; otherwise, it is labeled as low risk. The dataset includes 1,872 high-risk and 4,368 low-risk observations.

$$\mathit{\Delta} ROA_{i,t+1} = ROA_{i,t+1} - ROA_{i,t} \quad (1)$$

where $ROA_{i,t}$ denotes the return on assets of firm i in year t, and $\mathit{\Delta} ROA_{i,t+1}$ represents the one-year change in ROA from year t to year $t+1$.

$$Risk_{i,t+1} = 1\; if\; \mathit{\Delta} ROA_{i,t+1} \leq Q_{0.30}^{g(i),t};\; otherwise\; Risk_{i,t+1} = 0 \quad (2)$$

where $Risk_{i,t+1}$ is the binary label for profitability downside risk, $g(i)$ enotes the industry group of firm i, and $Q_{0.30}^{g(i),t}$, t is the 30th percentile of next-year ROA changes among firms in the same industry-year group. A firm is labeled as high risk when its next-year ROA change falls at or below this threshold.

DOI: 10.54254/2755-2721/2026.35021

The explanatory variables cover profitability, growth, liquidity, solvency, operating efficiency, and market and firm characteristics. Accounting variables are calculated from SEC Financial Statement Data Sets, while stock volatility is calculated from Stooq stock prices. The main indicators include ROA, operating margin, net profit margin, revenue growth, operating cash flow growth, current ratio, debt-to-asset ratio, asset turnover, stock volatility, firm size, and industry dummy variables. Table 1 summarizes key variables and expected economic meanings.

Table 1. Variable definitions and expected economic meanings

Variable Category	Variable	Measurement	Expected Effect	Economic Meaning
Profitability	ROA	Net income / total assets	Negative	Higher asset profitability indicates stronger operating performance.
Profitability	Operating margin	Operating income / revenue	Negative	Higher operating margin reflects better cost control and business efficiency.
Growth	Revenue growth	Change in revenue / lagged revenue	Negative	Slower revenue growth may indicate weakening demand or competitive pressure.
Growth	Operating cash flow growth	Change in operating cash flow / lagged operating cash flow	Negative	Cash flow deterioration may signal lower earnings quality.
Liquidity	Current ratio	Current assets / current liabilities	Negative	Higher liquidity reduces short-term financial pressure.
Solvency	Debt-to-asset ratio	Total liabilities / total assets	Positive	Higher leverage increases financial burden and risk exposure.
Efficiency	Asset turnover	Revenue / total assets	Negative	Higher asset turnover reflects better use of resources.
Market	Stock volatility	Standard deviation of monthly stock returns	Positive	Higher market volatility may reflect investor uncertainty about fundamentals.
Firm characteristic	Firm size	Natural logarithm of total assets	Ambiguous	Larger firms may be more stable but may also face organizational rigidity.

4. Methodology

The proposed framework links operating deterioration, financial pressure, and market uncertainty to future profitability decline. Figure 1 presents this mechanism rather than a purely technical workflow. In the empirical implementation, financial statement and market indicators are transformed into model features, the model estimates risk probabilities, and SHAP explains which economic channels drive the warning signal.

DOI: 10.54254/2755-2721/2026.35021

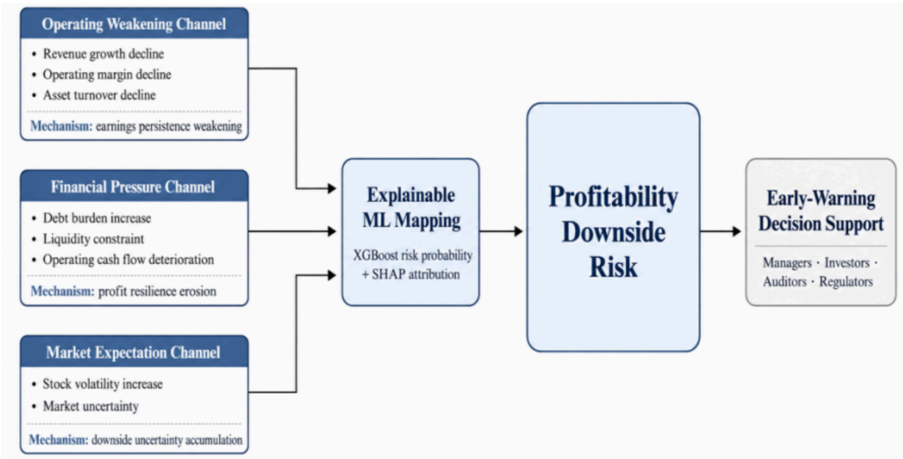

Figure 1. Conceptual mechanism of profitability downside risk early warning

Four models are compared. Logistic Regression is used as a transparent baseline. Random Forest captures threshold effects and variable interactions. XGBoost and LightGBM represent gradient boosting models suited to tabular data. Hyperparameters are selected on the validation set, and final evaluation is conducted on the out-of-sample test period.

$$\widehat{p}_{i,t+1} = f_\theta\left(X_{i,t}\right), \;\; \hat{y}_{i,t+1} = 1\left(\widehat{p}_{i,t+1} \geq \tau\right) \tag{3}$$

where $X_{i,t}$ denotes the vector of financial and market features for firm i in year t, $f_\theta(\cdot)$ is the trained prediction model with parameters θ, $\widehat{p}_{i,t+1}$ is the predicted probability of profitability downside risk, τ is the classification threshold, and $\hat{y}_{i,t+1}$ is the predicted risk label.

Model performance is assessed by Accuracy, Precision, Recall, F1-score, AUC, and PR-AUC. Accuracy is reported for completeness, but the early-warning task emphasizes Recall, F1-score, AUC, and PR-AUC. Missing a high-risk firm may be more costly than reviewing a false warning. The main classification threshold is 0.500. SHAP is applied to the XGBoost model for global and local interpretation. Global SHAP values rank features by mean absolute contribution, while local SHAP values explain why a specific firm is classified as high risk.

5. Empirical results and discussion

Table 2 reports out-of-sample test-set performance. Logistic Regression provides a reasonable baseline, with an AUC of 0.742 and an F1-score of 0.583.

Table 2. Model performance comparison on the test set

Model	Accuracy	Precision	Recall	F1-score	AUC	PR-AUC
Logistic Regression	0.711	0.563	0.604	0.583	0.742	0.526
Random Forest	0.758	0.628	0.671	0.649	0.807	0.614
XGBoost	0.781	0.662	0.704	0.682	0.836	0.653
LightGBM	0.775	0.651	0.698	0.674	0.829	0.641

DOI: 10.54254/2755-2721/2026.35021

The ensemble models outperform the linear baseline. XGBoost achieves the best results, with an Accuracy of 0.781, Recall of 0.704, F1-score of 0.682, AUC of 0.836, and PR-AUC of 0.653, while Random Forest and LightGBM also show competitive performance.

These results indicate that profitability downside risk is driven by multiple interacting signals rather than a single financial ratio. As shown in Figure 2, SHAP analysis further identifies revenue growth, operating margin, debt-to-asset ratio, ROA, operating cash flow growth, and stock volatility as the key drivers, reflecting demand weakening, cost pressure, financial pressure, and uncertainty.

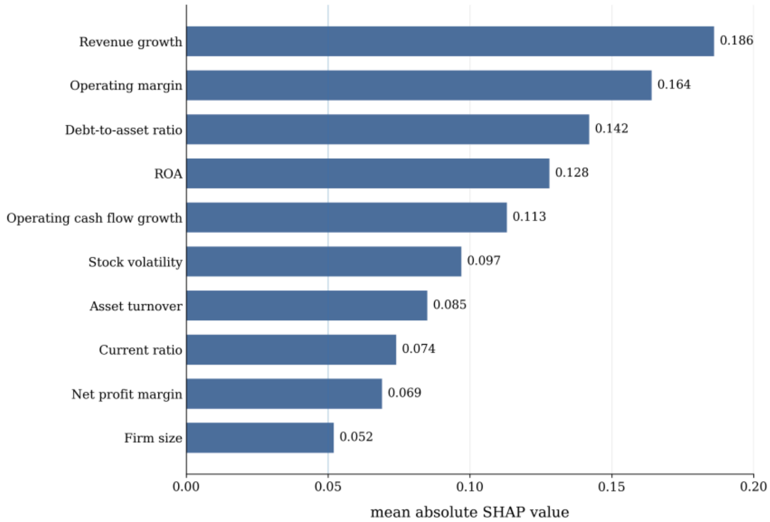

Figure 2. SHAP Summary of key drivers of profitability downside risk

6. Robustness checks

Three robustness checks support the stability of the main empirical results. Using next-year net profit margin decline as an alternative label still leaves XGBoost as the best model, with an AUC of 0.821 and an F1-score of 0.667. Lowering the warning threshold from 0.500 to 0.450 increases Recall from 0.704 to 0.746, with a moderate decline in Precision. Changing the time split to train on 2015–2020 and test on 2021–2024 produces similar performance, with an AUC of 0.828 and an F1-score of 0.675.

7. Conclusion

This paper proposes an explainable machine learning framework for early warning of profitability downside risk in listed companies. By predicting whether next-year ROA change falls into the lowest 30 percent within the same industry-year group, the framework shifts attention from late-stage financial failure to earlier operating deterioration. The empirical results show that XGBoost achieves the best predictive performance, while SHAP analysis identifies revenue growth, operating margin, debt-to-asset ratio, ROA, operating cash flow growth, and stock volatility as key risk drivers. These findings suggest that the proposed framework can improve risk screening efficiency and provide interpretable review signals for managers, investors, auditors, and regulators.

Future research can extend this framework using broader listed-company panels, textual disclosures, industry-specific models, and real-time monitoring systems.

DOI: 10.54254/2755-2721/2026.35021

References

[1] Beaver, W. H. (1966). Financial ratios as predictors of failure. Journal of Accounting Research, 4, 71-111. https: //doi.org/10.2307/2490171

[2] Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. The Journal of Finance, 23(4), 589-609. https: //doi.org/10.1111/j.1540-6261.1968.tb00843.x

[3] Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. Journal of Accounting Research, 18(1), 109-131. https: //doi.org/10.2307/2490395

[4] Campbell, J. Y., Hilscher, J., & Szilagyi, J. (2008). In search of distress risk. The Journal of Finance, 63(6), 2899-2939. https: //doi.org/10.1111/j.1540-6261.2008.01416.x

[5] Breiman, L. (2001). Random forests. Machine Learning, 45, 5-32. https: //doi.org/10.1023/A: 1010933404324

[6] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29(5), 1189-1232. https: //doi.org/10.1214/aos/1013203451

[7] Chen, T., & Guestrin, C. (2016). XGBoost: a scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794. https: //doi.org/10.1145/2939672.2939785

[8] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: a highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, 30.

[9] Shapley, L. S. (1953). A value for n-person games. In H. W. Kuhn & A. W. Tucker (Eds.), Contributions to the Theory of Games II (pp. 307-317). Princeton University Press.DOI:10.1515/9781400881970-018

[10] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30.