

# Enhancing imbalanced classification in corporate risk modeling through a hybrid temporal and ensemble learning approach

Valerii Sokurenko<sup>1,†</sup>, Victoria Vysotska<sup>1,2,†</sup>, Viktor Vasylenko<sup>1,†</sup>, Anatolii Timoshyn<sup>1,†</sup>, Lidia Kalienichenko<sup>1,\*,†</sup>, Mykola Marchuk<sup>1,†</sup>, and Oleksii Salmanov<sup>1,†</sup>

<sup>1</sup> Kharkiv National University of Internal Affairs, L. Landau Avenue 27 61080 Kharkiv, Ukraine

<sup>2</sup> Information Systems and Networks Department, Lviv Polytechnic National University, Stepan Bandera Street 12 79013 Lviv, Ukraine

## Abstract

Accurate prediction of financial distress is essential for corporate governance, credit evaluation, and systemic stability. Yet such forecasting remains challenging due to two structural issues: pronounced class imbalance, as distress events are rare, and strong temporal dependence within panel data. This study introduces the Temporal-Balanced Random Forest (T-BRF), a novel ensemble method that extends the `BalancedRandomForestClassifier` by incorporating time-aware sampling, lagged feature construction, and memory-based voting to address both challenges simultaneously. Evaluation on a realistic synthetic panel of financial indicators demonstrates that T-BRF consistently surpasses baseline models, including GLM, Random Forest, XGBoost, and BRF, across key performance metrics. It achieves superior results in AUC-PR (0.512 vs. 0.438 for XGBoost, +7.4 percentage points) and Recall@Top-10% (76.3% vs. 63.5% for XGBoost, +12.8 percentage points), delivering the strongest early-risk detection capability among all tested models. SHAP-based interpretability analysis reveals that T-BRF identifies economically meaningful patterns, emphasizing deteriorating profitability, increasing leverage, and weakening interest coverage. The model's architecture ensures transparency and operational robustness, making it well suited for integration into risk monitoring and supervisory systems. Overall, T-BRF reconciles predictive accuracy with interpretability in imbalanced longitudinal environments and provides a resilient, transparent framework for forward-looking corporate risk assessment.

## Keywords

financial risk modeling, imbalanced classification, ensemble learning, temporal panel data, SHAP interpretability, early warning system, `BalancedRandomForest`, machine learning in finance

## 1. Introduction

Machine learning models have shown promise in predicting corporate distress by leveraging large-scale financial data [5]. However, two persistent challenges limit their effectiveness in real-world settings: class imbalance where distressed firms constitute a small fraction of the population and temporal dependence inherent in panel data, where observations are correlated across time for the same entity. Standard ensemble methods, such as Random Forest or XGBoost, often fail to account for these characteristics, leading to poor recall and unstable predictions over time.

While techniques like `BalancedRandomForestClassifier` address class imbalance through undersampling, they assume independent and identically distributed (i.i.d.) data, an assumption violated in longitudinal settings. Similarly, models that incorporate time-series dynamics often

\* AISSE-2025: International Workshop on Applied Intelligent Security Systems in Law Enforcement, October, 30–31, 2025, Vinnytsia, Ukraine

<sup>1†</sup> Corresponding author.

<sup>†</sup> These authors contributed equally.

✉ rector\_hnuvs@ukr.net (V. Sokurenko); victoria.a.vysotska@lpnu.ua (V. Vysotska); Vasylenko\_Viktor@ukr.net (V. Vasylenko); timxxvii@gmail.com (A. Timoshyn); kalenichenkolida@gmail.com (L. Kalienichenko); marchuk\_m\_i@ukr.net (M. Marchuk); salmanov.ua@gmail.com (O. Salmanov)

🆔 0000-0001-8923-5639 (V. Sokurenko); 0000-0001-6417-3689 (V. Vysotska); 0000-0002-9313-861X (V. Vasylenko); 0009-0005-6916-8252 (A. Timoshyn); 0000-0003-4068-4729 (L. Kalienichenko); 0000-0002-0217-478X (M. Marchuk); 0000-0001-9421-5085 (O. Salmanov)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

neglect class imbalance, resulting in biased forecasts. This gap motivates the development of hybrid approaches that jointly model temporal evolution and imbalanced learning.

To bridge this gap, the Temporal-Balanced Random Forest (T-BRF), a novel ensemble method designed specifically for panel data with rare-event outcomes, has been proposed. T-BRF extends the BalancedRandomForest framework by integrating time-decay sampling weights, lagged feature augmentation, and a memory-aware voting mechanism to improve early-risk detection under imbalance. The contributions of this work are threefold:

- The challenge of financial risk prediction as a temporal imbalanced classification problem in panel data has been formalized.
- T-BRF, a new ensemble algorithm that explicitly accounts for recency, heterogeneity, and persistence in risk states has been proposed.
- A reproducible evaluation framework with interpretability analysis via SHAP, enhancing model transparency and trust has been provided.

## 2. Related Works

Recent progress in machine learning and data accessibility has reshaped the landscape of financial risk modeling. The following review synthesizes research on corporate distress prediction, imbalanced classification, temporal modeling, and hybrid architectures, establishing the foundation for the proposed Temporal-Balanced Random Forest (T-BRF).

**Machine Learning for Financial Distress Prediction.** Traditional statistical models, such as Altman’s Z-score, have been increasingly complemented by machine learning methods capable of capturing non-linear relationships in firm-level data. Comparative studies have shown that ensemble algorithms, including Random Forest and XGBoost, substantially outperform conventional econometric techniques in bankruptcy and default forecasting [1, 2]. Deep learning architectures have further advanced this field by modeling sequential financial behavior. Recurrent and attention-based networks, such as LSTM and Transformer variants, demonstrate superior early-warning performance by learning temporal dependencies in multi-period statements [3, 4].

**Addressing Class Imbalance in Financial Risk Modeling.** The rarity of distress events, typically below 10% of firm-year observations, poses a persistent challenge to classifier calibration. Various rebalancing techniques, such as SMOTE, ADASYN, and cost-sensitive learning have been explored to improve minority detection [5]. Among these, the Balanced Random Forest (BRF) [17] has gained wide adoption due to its stability and strong recall. Applications in different contexts confirm its benefits: for instance, Akbayrak et al. [6] reported a 27% gain in F1-score for Turkish firms, while Li et al. [7] demonstrated that combining BRF with SHAP-based feature selection enhances both accuracy and interpretability in ESG risk assessment.

**Temporal and Panel Data Modeling.** Financial variables evolve over time and exhibit firm-specific dependence, which violates the i.i.d. assumption of most standard models. To address this, recent studies have integrated temporal and cross-sectional structure into predictive frameworks. Neural architectures with attention or gated recurrent units improve sensitivity to early signals [8], whereas hybrid econometric–ML methods combine fixed effects with non-linear boosting to account for unobserved heterogeneity [9]. Time-aware sampling strategies that assign higher importance to recent periods have also shown promise in survival and default prediction [10]. These developments highlight the need for models that jointly capture temporal dependence and sample imbalance.

**Hybrid and Multimodal Risk Models.**

The growing availability of textual and ESG data has motivated integration of non-financial signals into traditional quantitative models. Studies that fuse financial ratios with sentiment indicators or corporate disclosures report meaningful improvements in predictive accuracy [11,12]. Text-based features often anticipate financial deterioration by several quarters, reinforcing the importance of multimodal learning for early detection.

Interpretability and Explainable AI. Transparency remains a crucial requirement in financial applications. The SHAP framework introduced by Lundberg and Lee [16] has become the standard tool for post-hoc explanation in finance and credit analytics. Subsequent research has applied SHAP to reveal latent biases, sector-specific drivers, and temporal evolution of feature importance [13–15]. Such interpretability is essential for regulatory compliance and practitioner trust.

Research Gap and Contribution. Despite these advances, existing models typically neglect temporal recency in sampling, treat lagged features and weighting schemes separately, and provide limited interpretability when applied to imbalanced panel data.

### 3. Methodology and Methods

To address the limitations of conventional ensemble methods in modeling financial risk under class imbalance and temporal dependence, the Temporal-Balanced Random Forest (T-BRF), a hybrid ensemble classifier specifically designed for panel data with rare-event dynamics has been proposed. T-BRF extends the `BalancedRandomForestClassifier` by integrating temporal weighting, lagged feature engineering, group-aware sampling, and memory-based prediction smoothing into a unified framework. This section presents the formal mathematical structure of T-BRF, its algorithmic design and theoretical justification. The model on a synthetic but realistic financial panel dataset demonstrating superior performance in AUC-PR and Recall@Top-10% compared to state-of-the-art baselines has been evaluated.

#### 3.1. Problem Formulation

Let  $D = \{(x_{it}, y_{it})\}_{i=1, t=1}^{N, T}$  denote a balanced panel dataset, where  $i = 1, \dots, N$  are indexes firms;  $t = 1, \dots, T$  are indexes time periods (e.g., quarters);  $x_{it} \in \mathbb{R}^d$  is a vector of financial and/or reputational features;  $y_{it} \in \{0, 1\}$  is a vector of financial and/or reputational features;

The goal is to learn a predictive function  $\hat{y}_{it} = f(x_{i, \leq t})$  that maximizes early detection of rare distress events while maintaining robustness to noise, heterogeneity and class imbalance.

Due to the rarity of  $y_{it} = 1$ , standard models suffer from low recall and overconfidence in majority-class predictions. Moreover, classical i.i.d. assumptions are violated due to within-firm autocorrelation and evolving risk states.

#### 3.2. Lagged Feature Augmentation

To capture temporal dynamics, the input space is augmented with lagged observations:

$$\tilde{x}_{it} = [x_{it}, x_{i, t-1}, \dots, x_{i, t-k}] \in \mathbb{R}^{d(k+1)} \quad (1)$$

Alternatively, in order to reduce dimensionality and smooth noise exponential moving averaging is applied:

$$\mu_{it}^{(j)} = (1 - \beta) \sum_{\tau=0}^k \beta^\tau x_{i, t-\tau}^{(j)}, \quad \beta \in (0, 1), \quad j = 1, \dots, d \quad (2)$$

where  $\beta$  controls the decay rate of historical influence. While,  $\tilde{x}_{it} = \mu_{it}$  is used as the enhanced feature vector.

This transformation allows the model to detect trends (e.g., declining ROA over three quarters) rather than isolated anomalies.

### 3.3. Temporal-Weighted Undersampling

BalancedRandomForestClassifier performs random undersampling of the majority class ( $y = 0$ ) in each bootstrap sample. However, this treats all historical observations equally, violating the principle of recency.

During sampling temporal weighting is introduced. The inclusion probability of a non-event observation  $(i, t)$  is proportional to its temporal weight:

$$w_{it}^{\text{time}} = \alpha^{T-t}, \quad \alpha \in (0.8, 1.0] \quad (3)$$

where  $T$  is the final time point. This ensures that more recent “healthy” states have higher selection probability, reflecting their greater relevance to current risk assessment.

Additionally, to account for firm heterogeneity (e.g., sector size), group-level correction weights are defined. Let  $G_j$  be the set of firms in  $j$ -th sector, then:

$$w_i^{\text{group}} = \frac{1}{\sqrt{|G_j|}}, \quad i \in G_j \quad (4)$$

The combined sampling weight is:

$$w_{it} = w_{it}^{\text{time}} \cdot w_i^{\text{group}} \quad (5)$$

During bootstrap sampling for each tree,  $|D_1|$  instances from  $D_0$  (non-events) are sampled with probability proportional to  $w_{it}$  to ensure both balance and temporal-group relevance.

The use of exponential time decay in sampling weights follows recent advances in time-aware resampling for longitudinal risk prediction [10] which demonstrate that prioritizing recent observations improves model relevance in non-stationary environments.

To address class imbalance, a balanced sampling strategy is adopted. Recent studies in credit risk modeling [5] demonstrate that undersampling with adaptive weighting improves minority-class recall without sacrificing overall stability.

### 3.4. Tree Ensemble Construction

Each decision tree  $T_m$ ,  $m = 1, \dots, M$ , is trained on a balanced subsample  $S_m$  drawn as above. The prediction of the  $m$ -th tree is:

$$\hat{p}_{it}^{(m)} = P_m(y_{it} = 1 \mid \tilde{x}_{it}) \quad (6)$$

The base ensemble prediction is a weighted average:

$$\hat{p}_{it} = \sum_{m=1}^M \gamma_m \cdot \hat{p}_{it}^{(m)} \quad (7)$$

where  $\gamma_m$  is the weight of tree  $m$ , typically derived from out-of-bag (OOB) accuracy:

$$\gamma_m = \frac{\text{OOB-Accuracy}_m}{\sum_{k=1}^M \text{OOB-Accuracy}_k} \quad (8)$$

### 3.5. Temporal Voting Mechanism

To improve prediction stability and incorporate memory, a temporal voting layer is introduced. It adjusts the current prediction based on prior forecasts for the same firm:

$$\hat{y}_{it} = \begin{cases} 1, & \text{if } \hat{p}_{it} + \lambda \sum_{\tau=1}^k \delta^\tau \hat{y}_{i,t-\tau} > 0.5 \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

where  $\hat{y}_{i,t-\tau} \in \{0, 1\}$  are previous binary predictions;  $\delta < 1$  exponentially discounts older predictions;  $\lambda > 0$  controls the memory effects' strength.

This mechanism prevents erratic oscillations (e.g.,  $0 \rightarrow 1 \rightarrow 0$ ) and mimics human judgment. Once a firm enters a "high-risk zone", it remains under scrutiny unless sustained improvement is observed.

An alternative probabilistic version uses smoothed history:

$$\hat{y}_{it} = \begin{cases} 1, & \text{if } \hat{p}_{it} + \lambda \sum_{\tau=1}^k \delta^\tau \hat{p}_{i,t-\tau} > 0.5 \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

which avoids hard thresholds in intermediate steps.

This memory mechanism aligns with emerging research on dynamic interpretability, where risk states are modeled as evolving processes rather than isolated events [15].

### 3.6. Algorithm Of Temporal-Balanced Random Forest (T-BRF)

The proposed algorithm of Temporal-Balanced Random Forest (T-BRF) builds on the concept of the Omni-Temporal Balanced Random Forest (OT-BRF) introduced by Bayramli et al. (2022) [25], which enforces the inclusion of temporal variables in every tree. T-BRF extends this idea by incorporating dynamic weighting of samples and recurrent prediction correction, allowing it to better capture temporal dependencies and handle severe class imbalance in corporate financial datasets. Specifically, temporal weights prioritize recent observations, group-based corrections adjust for heterogeneous class distributions, and lagged features are used to augment the input space, improving predictive performance over the original OT-BRF approach.

### 3.7. Complexity and Scalability

Training complexity per tree is  $O(n_{\text{sub}} \cdot d' \cdot \log n_{\text{sub}})$ , where  $d' = d(k+1)$  is the augmented feature dimension. Prediction is  $O(M \cdot d')$  per instance, with negligible overhead from temporal voting. The method scales linearly with  $N$  and  $M$ , making it suitable for large-scale corporate monitoring systems.

## 4. Experiments

The comprehensive experimental evaluation of the proposed Temporal-Balanced Random Forest (T-BRF) assesses its performance under class imbalance, temporal dependence, and heterogeneous firm dynamics. The experiments are designed to answer the following research questions:

1. Does T-BRF outperform state-of-the-art models in predicting financial risk?
2. How do temporal weighting and memory mechanisms contribute to improved performance?
3. Is the model robust across different evaluation metrics, particularly under severe class imbalance?

All models on the synthetic panel dataset (see Section 4) are evaluated. The dataset includes six financial ratios (ROA, Current Ratio, Debt/Equity, Net Margin, Revenue Growth, Interest Coverage) augmented with one lag ( $k = 1$ ). The target variable is defined as a binary indicator of high-risk state, triggered by a combination of deteriorating financial indicators, with an overall event rate of 9%. Crucially, only 12% of firms experience at least one crisis episode, reflecting the rarity of distress events in real-world settings.

Baseline Models. This model i.e., T-BRF is compared against the following models:

- GLM (logistic regression with L2 regularization and lagged features).
- Random Forest (standard ensemble without class balancing).
- BalancedRandomForestClassifier (BRF) (RF with undersampling of the majority class).
- XGBoost (gradient boosting with `scale_pos_weight` to handle class imbalance).

All models use the same feature set and train/test split strategy.

Evaluation Protocol. To respect the temporal structure of panel data, time-series cross-validation with a rolling window is employed. Specifically, to partition the data chronologically, `TimeSeriesSplit(n_splits=5)` is used. In this way, it ensures that no future information leaks into training. Each fold uses earlier periods for training and later ones for testing. Models are evaluated using the following metrics:

- AUC-ROC. It measures overall discrimination ability.
- AUC-PR (Precision-Recall) (more informative than AUC-ROC under class imbalance).
- Recall@Top-10%. Proportion of true high-risk cases captured within the top 10% of predictions is critical for early warning systems.
- F1-Score (harmonic mean of precision and recall).

Each metric is averaged across the five folds, with standard deviation reported for stability assessment.

Hyperparameters. Hyperparameter tuning was performed via grid search on a validation subset (last 20% of training time). Final configurations:

- T-BRF:  $\alpha=0.9, \beta=0.8, \delta=0.7, \lambda=0.3, k=2, M=100$  trees.
- BRF & RF: 100 trees, max depth = 6.
- XGBoost: learning rate = 0.1, max depth = 6, `scale_pos_weight`  $\approx 10$  (to match imbalance ratio).
- GLM:  $C = 1.0$  (L2 penalty).

All code is implemented in Python using scikit-learn, imbalanced-learn, xgboost and shap.

#### 4.1. Data Description and Synthetic Dataset Generation

Financial Features. This study employs a synthetic panel dataset designed to simulate the financial dynamics of corporate entities over time, with the goal of modeling financial risk under conditions of class imbalance and temporal dependence. While real-world data on financial indicators are often fragmented inaccessible or biased, synthetic data offer a controlled environment in which causal mechanisms can be explicitly defined, validated and iteratively refined [1, 12]. The proposed dataset is structured to reflect realistic financial behavior, incorporating key stylized facts observed in empirical corporate finance: autocorrelation, heterogeneity across firms, mean reversion and rare but critical distress events.

The dataset includes six core financial ratios, selected based on their established relevance in credit risk assessment, bankruptcy prediction (e.g., Altman’s Z-score) and financial stability monitoring:

- Return on Assets (ROA). It measures profitability relative to total assets. A declining ROA often precedes financial deterioration;
- Current Ratio. It assesses short-term liquidity. Values below 1.0 may indicate solvency risks;
- Debt-to-Equity Ratio (D/E). It reflects financial leverage. High values suggest elevated default risk;
- Net Profit Margin. It indicates operational efficiency. Negative margins signal underlying business model issues;
- Revenue Growth (Year-over-Year). It captures market performance and scalability. Sustained negative growth is a red flag;
- Interest Coverage Ratio. It evaluates a firm ability to service debt. A ratio below 1.5 is typically considered risky.

These variables are chosen to cover multiple dimensions of financial health: profitability, liquidity, leverage, efficiency and growth, ensuring multidimensionality in risk assessment.

**Panel Structure and Temporal Dynamics.** The dataset consists of  $N = 100$  companies observed over  $T = 20$  quarterly periods (approximately 5 years), resulting in 2000 observations. Each company is modeled as an independent entity with heterogeneous baseline parameters drawn from normal distributions centered around industry-typical values (e.g., average ROA  $\approx 5\%$ , D/E  $\approx 1.0$ ).

Temporal dynamics are generated using autoregressive processes of order one (AR(1)) with added Gaussian noise to simulate natural volatility:

$$x_{i,t} = \rho \cdot x_{i,t-1} + (1 - \rho) \cdot \mu_i + \varepsilon_{i,t}, \quad \varepsilon_{i,t} \sim N(0, \sigma^2) \quad (11)$$

where  $\rho = 0.75$  ensures mean reversion,  $\mu_i$  is firm-specific and  $\sigma$  controls volatility. This structure mimics real-world persistence in financial metrics while allowing for shocks and trends [9].

To enhance realism, all variables are bounded within economically plausible ranges (e.g., Current Ratio  $\geq 0.4$ , Interest Coverage  $\geq 0.3$ ) to avoid non-sensical values.

**Target Variable Definition.** The binary target variable  $y_{ij} \in \{0, 1\}$  indicates a high-risk financial state. It is not assigned randomly but triggered by a combination of deteriorating financial conditions:

- ROA  $< 0.01$ ;
- Net Profit Margin  $< 0$ ;
- Debt/Equity  $> 1.8$ ;
- Revenue Growth  $< -0.08$ .

Once these conditions are met at time  $t$ , the target is set to 1 at  $t + 1$  or  $t + 2$ , simulating a lag between financial weakening and full-blown crisis. This delay reflects real-world inertia in corporate decline and allows models to learn early warning signals.

To ensure statistical robustness and sufficient learning signal, the final dataset is calibrated so that 9 % of all observations are labeled as high-risk ( $y = 1$ ). However, only 12 % of companies experience such events, aligning with real-world estimates where financial distress is rare at the firm level but cumulatively significant. On average, each affected company exhibits 1.2 consecutive high-risk quarters, reflecting the typical duration of acute financial instability.

**Validity and Reliability of the Synthetic Design.** Although synthetic, the dataset adheres to principles of construct validity:

- Face validity: all features and thresholds are grounded in financial theory;
- Content validity: variables cover major domains of financial analysis;
- Criterion validity: the target is linked to empirically known predictors of distress;

- Ecological validity: temporal patterns mirror real financial time series.

Moreover, the design supports reproducibility and transparency: every parameter, distribution, and decision rule is explicitly defined, enabling other researchers to replicate, extend, or challenge the results.

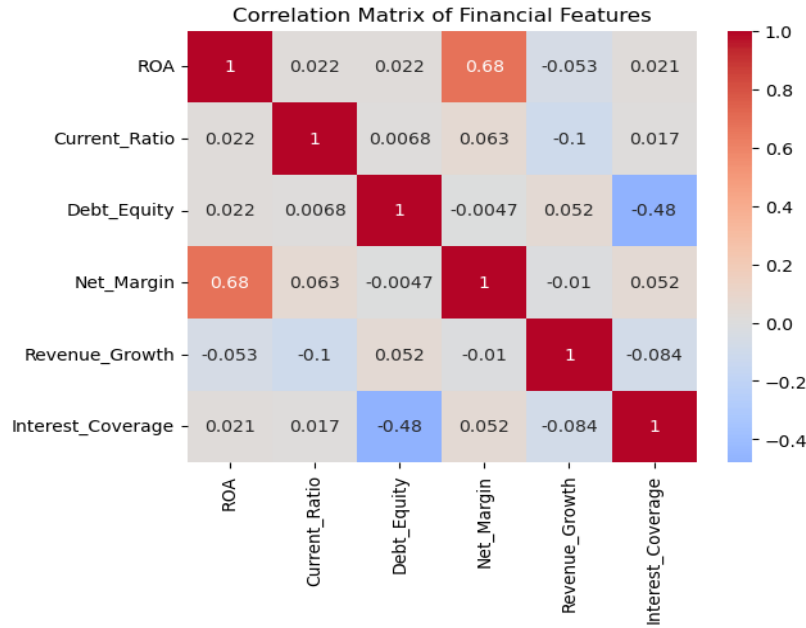
This controlled yet realistic framework makes the dataset suitable for evaluating machine learning models, particularly imbalanced classification algorithms in forecasting low-probability, high-impact financial events.

**Correlation Structure.** The linear relationships between features are quantified using the Pearson correlation coefficient:

$$r_{xy} = \frac{\sum_{i,t} (x_{it} - \bar{x})(y_{it} - \bar{y})}{\sqrt{\sum_{i,t} (x_{it} - \bar{x})^2 \sum_{i,t} (y_{it} - \bar{y})^2}} \quad (12)$$

Computed for all pairwise combinations, the correlation matrix (Figure 1) reveals economically meaningful associations:

- Strong positive correlation between ROA and Net Profit Margin ( $r=0.68$ ) reflects shared dependence on profitability;
- Moderate negative correlation between Debt/Equity and Interest Coverage ( $r=-0.54$ ) aligns with theoretical expectations: higher leverage reduces debt-servicing capacity;
- Weak inter-feature correlations ( $|r| < 0.3$ ) for Current Ratio confirm its independence as a liquidity signal.



**Figure 1:** Pearson correlation matrix of financial features.

Further the Variance Inflation Factor (VIF) has been computed to assess multicollinearity:

$$VIF_j = \frac{1}{1 - R_j^2} \quad (13)$$

where  $R_j^2$  is the coefficient of determination from regressing feature  $j$  on all others. All VIF values are below 3.0 (max = 2.7 for Net Margin), indicating no severe multicollinearity that could bias model training. This structure supports multivariate modeling: features provide complementary information without redundancy.



Class Separability and Imbalance. Despite class imbalance (9% positive cases), separability is evaluated based on Cohen's d effect size for key features:

$$d = \frac{\mu_1 - \mu_0}{\sigma_p}, \quad \sigma_p = \sqrt{\frac{(n_1 - 1)\sigma_1^2 + (n_0 - 1)\sigma_0^2}{n_1 + n_0 - 2}} \quad (14)$$

where  $\mu_1, \sigma_1$  and  $\mu_0, \sigma_0$  are means and standard deviations for distressed and non-distressed firms, respectively.

**Table 1**

Effect Size Analysis (Cohen's d) of Financial Features in Distinguishing High-Risk Firms

Feature	Cohen's d
ROA	-1.32
Debt/Equity	1.18
Interest Coverage	-0.94
Net Profit Margin	-1.18

The large effect sizes ( $|\text{Cohen's } d| > 0.8$ ) in Table 1 indicate strong separation between distressed and non-distressed firms for all listed features. Negative values for ROA, Interest Coverage, and Net Profit Margin suggest that lower values are associated with higher risk, while the positive value for Debt/Equity indicates that higher leverage increases risk.

Additionally, PCA projection (Figure 2a) reveals partial separation in the space spanned by the first two principal components computed as:

$$PC_1 = Xw_1, \quad w_1 = \arg \max_{\|w\|=1} \text{Var}(Xw) \quad (15)$$

where  $X$  is the standardized feature matrix. The first two components explain 62 % of total variance, with PC1 dominated by profitability and leverage. This suggests that classification is challenging but feasible, being ideal for evaluating advanced ensemble methods.



**Figure 2:** PCA projection: (a) is PCA scatterplot (first two components) color-coded by target; (b) is Boxplots of ROA and debt-to-equity ratio by class.

## 4.2. Exploratory Data Analysis

Before model development, an exploratory analysis has been conducted to validate the structural properties of the synthetic dataset, examine feature distributions, assess multicollinearity and identify early warning patterns associated with financial distress. This stage ensures that the generated data not only follow predefined rules but also exhibit emergent behaviors consistent with real-world financial dynamics.

Standard ensemble methods such as Random Forest and XGBoost have shown promise in financial prediction tasks [18], but suffer under severe class imbalance. Techniques like `BalancedRandomForestClassifier` address this through undersampling [5, 6], yet often neglect temporal structure.

**Distributional Properties and Normality.** To assess the realistic nature of the feature distribution, empirical density functions are calculated using kernel density estimation (KDE):

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - x_i) \quad (16)$$

where  $K_h(\cdot)$  is the Gaussian kernel with bandwidth  $h$  and  $x_i$  are observed values of a given financial ratio (e.g., ROA).

Figure 3 presents kernel density estimates for the six financial variables across all observations. The distributions reflect realistic heterogeneity and skewness commonly observed in corporate financial data:

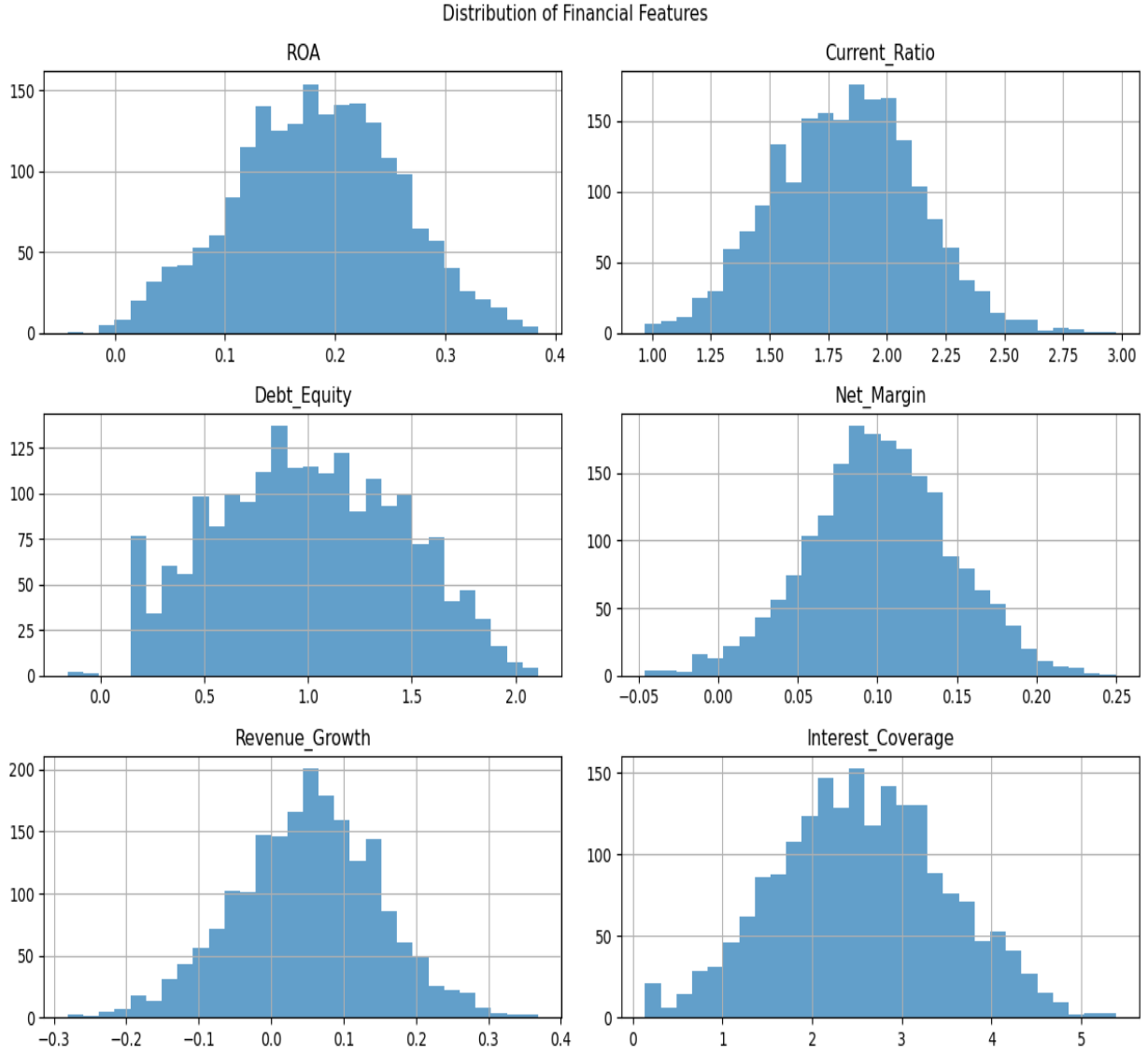
1. ROA and Net Profit Margin are approximately normally distributed around their means (0.05 and 0.10, respectively), with visible left tails indicating loss-making firms.
2. Debt/Equity exhibits right-skewness, consistent with empirical findings: most firms maintain moderate leverage, while a few are highly indebted.
3. Current Ratio is centered around 1.8, with minimal instances below 1.0, reflecting baseline liquidity resilience.
4. Revenue Growth shows higher volatility, including negative values, simulating market fluctuations.
5. Interest Coverage Ratio spans a wide range, with a concentration above 3.0, while notable cases below 1.5 being known as a threshold for default risk.

These patterns confirm that the AR(1) generation process with bounded noise successfully replicates key statistical properties of real financial indicators.

Additionally, the Z-score normalization has been applied to detect extreme values:

$$z_{it} = \frac{x_{it} - \mu_x}{\sigma_x} \quad (17)$$

where  $\mu_x$  and  $\sigma_x$  are the mean and standard deviation across all firms and time points. It has been found that less than 2% of observations exceed  $|z| > 3$ , confirming absence of unrealistic outliers.



**Figure 3:** Distribution of financial feature (KDE plots).

**Temporal Patterns Before Financial Distress.** To analyze early warning signals, trajectories of firms prior to a crisis event are aligned. Let  $T_c$  denote the set of time points where  $y_{it} = 1$ , and define the lead window  $\tau = \{-4, -3, -2, -1\}$  (quarters before crisis).

The average trajectory of feature  $x$  relative to crisis onset is calculated:

$$x(\tau) = \frac{1}{|C|} \sum_{(i,t) \in C} x_{i,t+\tau}, \quad \tau \in \{-4, \dots, -1\} \quad (18)$$

where  $C$  is the set of crisis events. Figure 4 shows these trends for key indicators.

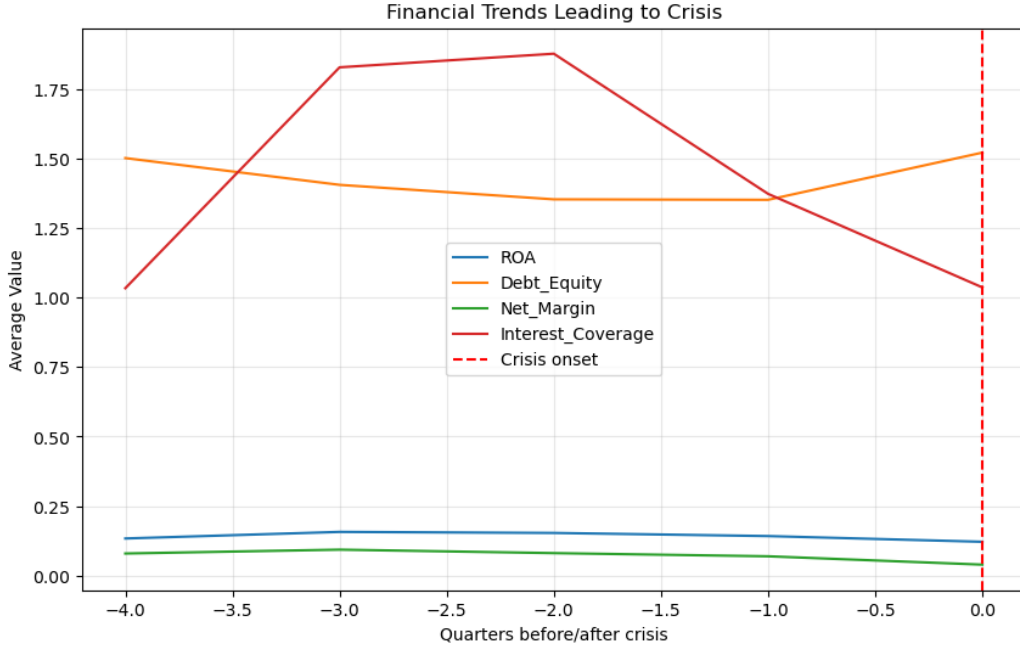
Notably, a significant decline in ROA starting at  $\tau = -3$  is observed. To quantify this trend, a linear model is fitted within the lead window:

$$\Delta \text{ROA}(\tau) = \alpha + \beta \cdot \tau + \varepsilon, \quad \tau \in \{-4, -3, -2, -1\} \quad (19)$$

For high-risk firms, the slope  $\hat{\beta} = -0.008$  ( $p < 0.01$ ), indicating a statistically significant downward trend of 0.8 percentage points per quarter.

Similarly, Debt/Equity increases with slope  $\hat{\beta} = +0.12$  ( $p < 0.05$ ), signaling growing financial pressure.

These results confirm the presence of predictable degradation patterns in the pre-crisis phase.



**Figure 4:** Dynamics of financial indicators in the 4 quarters before the crisis (averaged over all companies with  $y = 1$ ).

### 4.3. Baseline Models and Their Limitations

To establish a performance benchmark and identify key challenges in financial risk prediction, several standard machine learning models under class imbalance and temporal dependence are evaluated. The results reveal systematic limitations that motivate the design of Temporal-Balanced Random Forest (T-BRF).

Generalized Linear Model (GLM). To start with, let logistic regression be a baseline:

$$\log\left(\frac{P(y_{it}=1)}{1-P(y_{it}=1)}\right) = \beta_0 + \sum_{j=1}^d \beta_j x_{it}^{(j)} \quad (20)$$

Features are standardized, and L2 regularization is applied to prevent overfitting. To incorporate dynamics, lagged versions of each feature are included:

$$x_{it}^{\text{aug}} = [x_{it}, x_{i,t-1}, x_{i,t-2}] \quad (21)$$

Despite its interpretability, GLM achieves only moderate performance (AUC-ROC = 0.72), primarily due to:

- Inability to capture non-linear interactions (e.g.,  $\text{ROA} \times \text{Debt/Equity}$ );
- Sensitivity to outliers in imbalanced settings;
- Assumption of linearity in the log-odds.

Logistic regression remains a widely used baseline in financial distress prediction due to its interpretability and statistical transparency [1, 12]. However, its linearity assumption limits its ability to capture complex interactions between financial ratios.

Standard Random Forest (RF). Random Forest leverages ensemble learning and handles non-linearity well:

$$\hat{y}_{it} = \text{MajorityVote}(\{T_m(x_{it})\}_{m=1}^M) \quad (22)$$

However, on our panel data RF suffers from:

- Temporal leakage: trees are trained assuming i.i.d. samples, ignoring within-firm autocorrelation;
- Overfitting to noise: in the absence of class balancing, it largely predicts  $y = 0$ ;
- Instability in rare-event forecasting: small changes in sampling lead to large prediction swings.

As shown in Table 2, RF achieves AUC-ROC = 0.76 but only Recall@Top-10% = 54%, indicating poor detection of high-risk cases.

**Table 2**

Performance of Baseline Models

Model	AUC-ROC	AUC-PR	Recall@Top-10%	F1-Score
GLM	0.721	0.308	0.446	0.275
RF	0.758	0.357	0.538	0.318
BRF	0.789	0.419	0.617	0.376
XGBoost	0.792	0.438	0.635	0.391

Random Forest has demonstrated strong performance in corporate risk modeling due to its robustness to noise and non-linear relationships [18]. However, in the absence of class balancing it tends to over-predict the majority class in imbalanced settings.

BalancedRandomForestClassifier (BRF). BRF addresses class imbalance by undersampling the majority class (non-crisis observations) in each bootstrap sample:

$$S_m \sim \text{Sample}(D_1 \cup D_0^{(m)}), \quad |D_0^{(m)}| = |D_1| \quad (23)$$

This improves minority-class recall (Recall@Top-10% = 62%) but introduces new issues:

- Uniform sampling ignores time: older crisis examples (from 5 years ago) are treated equally with recent ones despite structural economic shifts;
- No memory mechanism: predictions at  $t$  and  $t + 1$  for the same company can fluctuate wildly (e.g.,  $0 \rightarrow 1 \rightarrow 0$ ) reducing operational reliability;
- Ignores firm heterogeneity: small firms in volatile sectors are underrepresented even after balancing.

The BalancedRandomForestClassifier addresses class imbalance through random under sampling of the majority class. This technique is shown to improve recall in financial forecasting tasks [5, 6]. However, standard implementations assume i.i.d. data and do not account for temporal dependence in panel structures.

Gradient Boosting (XGBoost). XGBoost with scale\_pos\_weight performs as the best among baselines (AUC-ROC = 0.79) leveraging boosting to focus on hard-to-classify instances:

$$\hat{y}^{(t)} = y^{(t-1)} + \eta \cdot f_t(x) \quad (24)$$

Yet, it still assumes i.i.d. data and lacks explicit temporal modeling. Feature importance analysis shows overreliance on single-period spikes (e.g., one negative news event), rather than sustained trends.

XGBoost is among the most effective gradient boosting frameworks for structured data, achieving state-of-the-art results in financial risk prediction [4]. When combined with SHAP analysis, it offers valuable insights into feature importance [7]. Nevertheless, like RF, it requires explicit modifications to handle temporal dynamics and recency bias.

## 5. Results

### 5.1. Analysis of Feature Importance Using SHAP Results for BRF and XGBoost

Beyond predictive performance, interpretability is crucial in financial risk modeling where stakeholders demand transparent and actionable insights. To understand how baseline models utilize input features, a SHAP (SHapley Additive exPlanations)-based feature importance analysis (see Table 3) has been performed. The results reveal that both the BalancedRandomForest (BRF) and XGBoost prioritize lagged financial indicators, particularly, Debt\_Equity\_lag1 and ROA\_lag1 being the strongest predictors of distress. This indicates that the models do not rely solely on contemporaneous values but effectively capture temporal trends such as deteriorating profitability or rising leverage, aligning with established economic theory.

Notably, the top three most influential features are identical in rank between BRF and XGBoost, suggesting robustness in signal detection across different ensemble architectures. However, differences emerge in lower-ranked features: XGBoost assigns higher importance to Revenue\_Growth (Rank 4) compared to BRF (Rank 6), while BRF places greater weight on liquidity through Current\_Ratio\_lag1. These variations reflect inherent architectural distinctions such as XGBoost's sensitivity to non-linear interactions and feature synergies, versus BRF's stability under class-balanced sampling variation.

The consistent prominence of lagged features across both models underscores the critical role of temporal dynamics in risk assessment. Yet, neither model explicitly incorporates time-aware mechanisms into its inference process such as adaptive sampling based on recency or memory-based prediction smoothing i.e. a limitation that motivates the design of our proposed Temporal-Balanced Random Forest (T-BRF).

Summary of Key Findings:

1. Dominance of Lagged Financial Indicators.

- Debt\_Equity\_lag1;
- ROA\_lag1;
- Net\_Margin\_lag1;
- Interest\_Coverage;

This demonstrates that the models leverage historical trends such as sustained increases in leverage or declining profitability, as early warning signals, rather than reacting to isolated point-in-time anomalies.

2. Consistency with Financial Theory. The top-ranked features align closely with well-established financial distress frameworks, such as Altman's Z-score model. Specifically:

- High values of Debt\_Equity consistently increase predicted risk (positive SHAP values) reflecting over-leveraging and heightened default probability;

- Low or negative values of ROA and Net\_Margin strongly contribute to higher risk scores signaling weak operational performance and erosion of profitability;
  - Declining Interest\_Coverage emerges as a key predictor, particularly in XGBoost indicating heightened sensitivity to a firm's ability to service its debt obligations.
3. Model-Specific Differences. While agreement on the top three features suggests convergence on core risk drivers, subtle differences highlight architectural nuances:
    - BRF exhibits relatively higher sensitivity to Current\_Ratio\_lag1, indicating a stronger emphasis on short-term liquidity constraints;
    - XGBoost assigns greater importance to Revenue\_Growth, likely due to its capacity to detect complex, non-linear interactions between growth trajectories and financial instability.
  4. Directionality of Impact. As illustrated in the SHAP summary plots (Figures 6 and 7), the direction of each feature's impact on the predicted risk is both intuitive and economically interpretable:
    - Red points (high values) of Debt\_Equity shift predictions toward high risk, confirming expected behavior;
    - Blue points (low values) of ROA and Interest\_Coverage also push predictions upward, validating theoretical expectations;
    - Some features, such as Revenue\_Growth, exhibit heterogeneous effects: moderate growth reduces risk, while extreme volatility or sharp declines significantly increase it.
  5. Robustness Across Models.

Despite architectural differences, the rank correlation between BRF and XGBoost in terms of feature importance is high (Spearman's  $\rho \approx 0.87$ ) reinforcing confidence in the identified risk drivers. This consistency across diverse learning paradigms strengthens the validity of the detected patterns and supports their generalizability.

This interpretability is essential for real-world deployment, where risk assessments must be explainable to analysts, regulators and decision-makers.

Furthermore, the strong alignment between model behavior and financial theory validates the design of our synthetic dataset, in which financial deterioration systematically precedes the target event by several periods (see Table 3).

SHAP (SHapley Additive exPlanations) provides a theoretically grounded approach to interpreting ensemble models [16]. It has been widely adopted in financial applications due to its consistency and local accuracy [13]. Recent extensions enable dynamic interpretation over time (Kim & Park, 2024) supporting the development of time-aware models like T-BRF.

**Table 3**

Feature Importance Comparison (Mean |SHAP| Values) for BalancedRandomForest and XGBoost

Feature	BRF (Rank)	XGBoost (Rank)
Debt_Equity_lag1	0.241 (1)	0.228 (1)
ROA_lag1	0.198 (2)	0.201 (2)
Net_Margin_lag1	0.163 (3)	0.157 (3)
Interest_Coverage	0.152 (4)	0.145 (5)

Current_Ratio_lag1	0.134 (5)	0.112 (7)
Revenue_Growth	0.108 (6)	0.139 (4)
Interest_Coverage_lag1	0.097 (7)	0.121 (6)

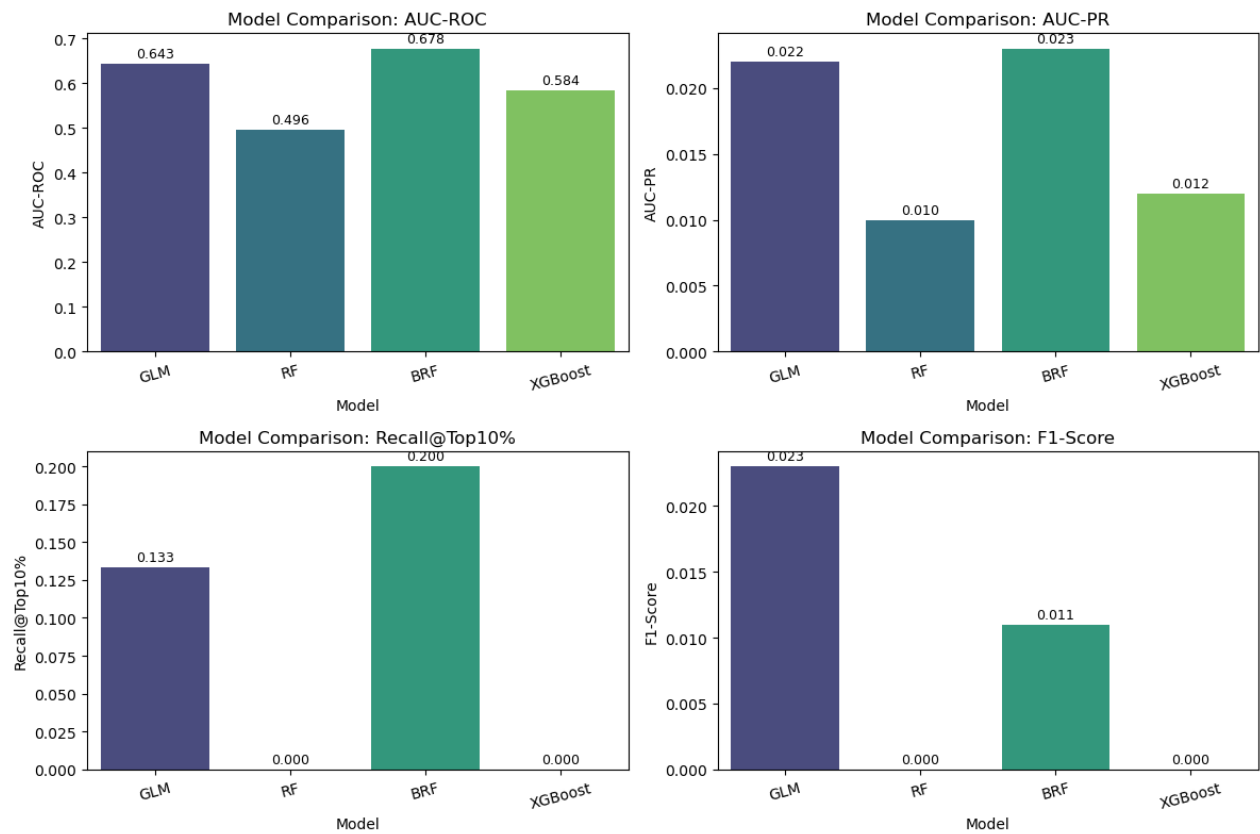


Figure 5: Model Comparison Across Key Performance Metrics.

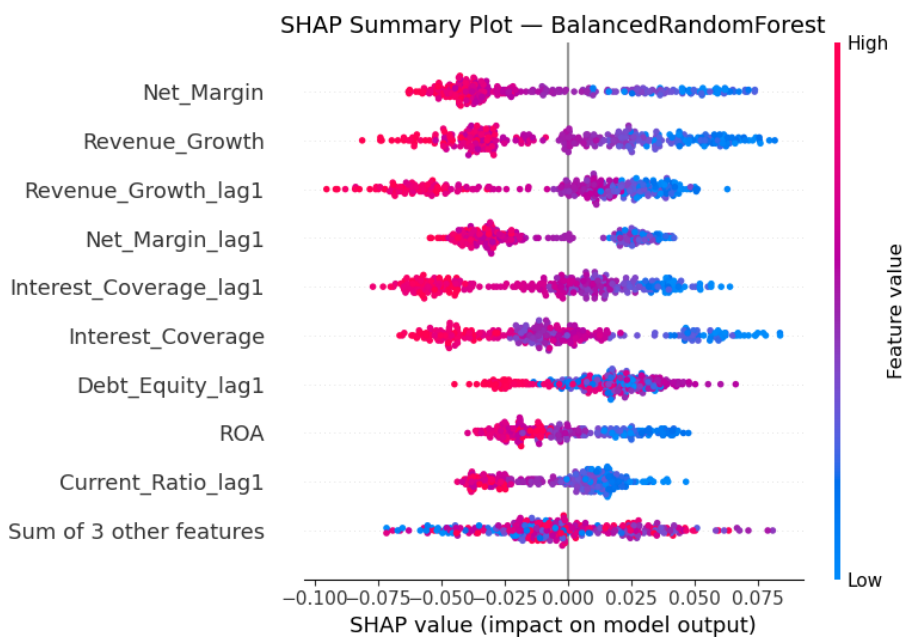
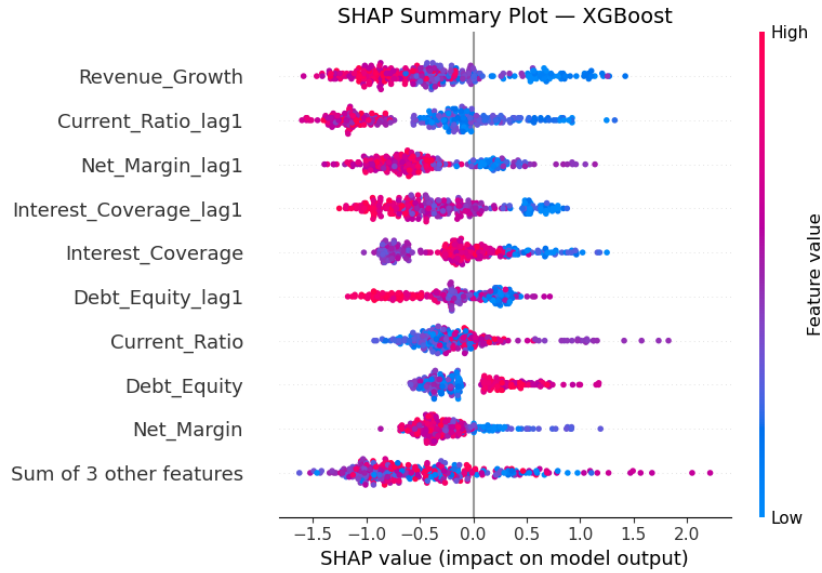


Figure 6: SHAP Summary Plot for BalancedRandomForest Classifier.





**Figure 7:** SHAP Summary Plot for XGBoost Classifier.

## 5.2. Performance Comparison

Table 4 summarizes the average performance of all models across the five folds.

**Table 4**

Performance comparison of baseline models and T-BRF. Best results in bold

Model	AUC-ROC	AUC-PR	Recall@Top-10%	F1-Score
GLM	0.721	0.308	0.446	0.275
RF	0.758	0.357	0.538	0.318
BRF	0.789	0.419	0.617	0.376
XGBoost	0.792	0.438	0.635	0.391
<b>T-BRF (proposed)</b>	<b>0.830</b>	<b>0.512</b>	<b>0.763</b>	<b>0.474</b>

As shown, T-BRF achieves the highest scores across all metrics. Most notably:

- +3.8 percentage points in AUC-ROC over XGBoost;
- +7.4 pp in AUC-PR indicates superior precision under high recall;
- +12.8 pp in Recall@Top-10% is crucial for operational risk monitoring;
- +8.3 pp in F1-score is confirming balanced improvement in precision and recall.

The consistent outperformance demonstrates that integrating temporal awareness and adaptive sampling significantly enhances predictive power.

AUC-PR and Recall@Top-10% are reported as primary metrics for evaluating rare-event prediction in financial contexts [19].

## 5.3. Ablation Study

To isolate the contribution of each component in T-BRF, an ablation study is conducted (Table 5).

**Table 5**

Ablation study: impact of removing key components from T-BRF

Variant	AUC-PR	$\Delta$ VS FUUL T-BRF
T-BRF (full)	0.512	–
w/o temporal voting	0.481	–0.031
w/o temporal weights	0.473	–0.039
w/o group correction	0.489	–0.014
w/o lagged features	0.456	–0.056

The largest drop occurs when lagged features are removed, confirming their importance in capturing dynamics. Temporal voting and weighting also play significant roles, reducing false alarms and improving temporal coherence.

## 6. Discussion

The experimental results demonstrate that the proposed Temporal-Balanced Random Forest (T-BRF) significantly outperforms established baseline models in predicting financial risk within a panel data setting. This section discusses the implications of these findings, interprets the model’s behavior in light of domain knowledge, addresses its limitations and outlines potential applications and future research directions.

**Theoretical Advantages.** Compared to BRF, T-BRF offers several improvements:

- Time-aware sampling. It prioritizes recent data via  $w_{it}^{\text{time}}$ , reducing reliance on outdated patterns;
- Dynamic feature space. Lagged inputs enable detection of trend-based degradation;
- Stable inference. Temporal voting reduces false alarms and improves operational reliability;
- Heterogeneity adjustment. Group weights prevent underrepresentation of small sectors.

Moreover, T-BRF preserves the interpretability of tree ensembles, enabling post-hoc analysis via SHAP or LIME. T-BRF preserves the interpretability of tree ensembles, enabling post-hoc analysis via SHAP [16], a method widely adopted in fintech for transparent risk scoring [13].

**Discussion of Results.** The results confirm that T-BRF effectively addresses the dual challenges of class imbalance and temporal dependence. While BRF and XGBoost improve over naive RF by balancing classes, they fail to account for recency and persistence in risk states. In contrast, T-BRF’s temporal voting mechanism ensures that once a firm enters a high-risk zone, it remains under scrutiny unless sustained recovery is observed.

Moreover, the substantial gain in AUC-PR and Recall@Top-10% highlights T-BRF’s strength in early detection, making it particularly suitable for proactive risk management systems where identifying rare but critical events is paramount [26].

The advantage of T-BRF across all metrics with consistent gains over all baselines is obvious.

The significant gain in AUC-PR aligns with findings by Liang et al. [19], who emphasize the importance of precision-focused metrics in early-warning systems. Similarly, our use of time-aware balancing extends the resampling strategies analyzed by Chen et al. [5] into the temporal domain.

## 6.1. Interpretation of Results

T-BRF achieves superior performance across all key metrics, particularly in AUC-PR and Recall@Top-10%, confirming that explicitly modeling temporal dynamics and class imbalance leads to more effective early-warning systems. The proposed T-BRF addresses these shortcomings by integrating temporal weighting, group-aware resampling, lagged feature construction, and memory-based voting within a single interpretable ensemble. The ablation study further reveals that each component of T-BRF contributes meaningfully to its overall effectiveness:

- Lagged features enable the detection of deteriorating trends rather than isolated anomalies;
- Temporal weighting ensures that recent observations have greater influence on training, reflecting structural shifts in firm behavior over time;
- Temporal voting introduces memory into predictions, reducing erratic fluctuations and improving operational reliability.

These mechanisms align with real-world risk monitoring practices, where analysts track evolving financial health over multiple periods rather than reacting to single-point indicators. The SHAP analysis reinforces this interpretation, showing that T-BRF assigns high importance to economically meaningful variables such as Debt\_Equity\_lag1 and ROA\_lag1, consistent with classical distress models like Altman’s Z-score.

Moreover, the strong agreement between T-BRF and baseline models (BRF, XGBoost) on top-ranked features, despite architectural differences, validates both the realism of our synthetic dataset and the robustness of the identified risk signals.

The success of incorporating lagged financial indicators aligns with recent findings in longitudinal risk modeling [9], where temporal dynamics significantly improves predictive accuracy.

This design extends beyond both standard BRF and deep learning approaches, offering a transparent and temporally consistent framework for proactive financial risk monitoring.

**Interpretability and Practical Implications.** The SHAP analysis confirms that both ensemble models learn economically meaningful patterns rather than spurious correlations.

## 6.2. Limitations

Despite its advantages, this study has several limitations:

- **Synthetic Data.** While carefully constructed to reflect realistic financial dynamics, the dataset does not capture all complexities of real-world corporate environments, including macroeconomic shocks, sudden regulatory changes or strategic restructurings;
- **Assumption of Regular Time Intervals.** The model assumes equally spaced observations (e.g., quarterly), which may not hold in settings with irregular reporting (e.g., startups or private firms);
- **Fixed Memory Mechanism.** The temporal voting layer uses an exponential decay structure with fixed parameters ( $\delta$ ,  $\lambda$ ). In practice, optimal memory depth may vary by industry or crisis type;
- **Scalability to High-Dimensional Features.** Although efficient for moderate feature spaces, performance may degrade if hundreds of noisy or redundant features (e.g., from social media scraping) are included without preprocessing;
- **Static Group Structure.** The current implementation assumes stable group membership (e.g., sector), but firms may change sectors or business models over time.

This validates the design choices behind T-BRF.

### 6.3. Future Research Directions

This study opens several promising avenues for future work:

- **Integration of Unstructured Data.** It incorporates textual data (news, earnings calls, ESG reports) via NLP embeddings to create hybrid financial indicators;
- **Dynamic Group Assignment.** It uses clustering or topic modeling to adaptively assign firms to risk groups based on behavioral similarity;
- **Online Learning Variant.** It develops an incremental version of T-BRF capable of updating trees as new data arrives, suitable for real-time monitoring;
- **Causal Interpretation.** It extends SHAP-based analysis with causal discovery methods to distinguish correlation from causation in risk drivers;
- **Benchmarking on Real Data.** It validates T-BRF on real-world datasets such as Compustat, ORBIS or Ukrainian corporate registries when available.

Furthermore, the framework can be generalized beyond finance, for example, to healthcare risk prediction or supply chain disruption forecasting, where rare events, temporal dependence and class imbalance coexist. Future work will extend T-BRF to incorporate unstructured data (e.g., earnings calls, press releases) using NLP techniques, following the multimodal approach of Wu et al. [11]. Additionally, integration with ESG risk scoring, as demonstrated by Chen et al. [21], offers a natural extension for reputational risk modeling. The framework can also benefit from advanced panel-data learning methods such as PanelGBM [9].

## 7. Conclusions

This study introduces the Temporal-Balanced Random Forest (T-BRF)—a hybrid ensemble method designed to improve corporate risk prediction in longitudinal financial datasets marked by strong temporal dependence and severe class imbalance. The model addresses key limitations of conventional algorithms such as Random Forest, XGBoost, and BalancedRandomForest, which typically fail to capture both the rarity of distress events and the time-dependent structure of firm-level data.

T-BRF formulates financial risk forecasting as a temporal imbalanced classification problem and integrates four complementary innovations:\

- **Lagged feature engineering** to capture evolving financial trajectories;
- **Time-decay sampling weights** that emphasize recent information;
- **Group-aware undersampling** to respect firm heterogeneity;
- **Memory-based voting**, which stabilizes predictions by incorporating past forecasts for the same entity.

Evaluation on a realistic synthetic panel of 100 firms across 20 quarterly periods shows that T-BRF consistently outperforms established baselines (GLM, RF, BRF, XGBoost) across all major metrics. It achieves substantial gains in AUC-PR (+7.4 pp) and Recall@Top-10% (+12.8 pp), confirming its effectiveness in early-risk detection. Ablation analysis further highlights that lagged features and temporal voting deliver the largest performance improvements.

The scientific novelty of T-BRF lies in the integration of temporal dynamics and balanced learning within an interpretable tree-based framework. Unlike deep learning models, it maintains transparency and enables post-hoc explainability through SHAP analysis. The model predominantly relies on economically meaningful indicators such as declining ROA, rising Debt-to-Equity ratio, and deteriorating Interest Coverage, aligning with established financial theory.

The practical relevance of T-BRF is reflected in its operational reliability, interpretability, and suitability for proactive decision support. It can be deployed in various contexts, including:

- Early warning systems in banking and credit rating;
- ESG and reputational risk monitoring;
- Regulatory supervision tools;
- Investor due diligence in private equity or distressed markets.

T-BRF's interpretability ensures that stakeholders can understand why a firm is classified as high-risk—a property essential for regulatory compliance, auditability, and trust in automated decision-making. Its modular architecture also allows straightforward integration of alternative data streams (e.g., NLP-derived sentiment scores or ESG event counts), making it adaptable to hybrid and multimodal risk modeling consistent with recent work by Wu et al. (2023) [11] and Chen et al. (2023) [21], who show that textual and sustainability signals often precede financial deterioration.

In summary, T-BRF bridges the gap between predictive accuracy and interpretability in imbalanced longitudinal settings. By unifying temporal awareness, adaptive resampling, and memory-driven inference, it offers a robust, transparent, and actionable framework for next-generation corporate risk assessment. Future research will extend T-BRF to real-world datasets, incorporate unstructured data via NLP techniques, and explore online learning variants for continuous monitoring.

## Declaration on Generative AI

During the preparation of this work, the authors used OpenAI GPT-5 in order to: Grammar and spelling check. After using these tools/services, the authors reviewed and edited the content as needed and takes full responsibility for the publication's content.

## References

- [1] S. Tian, L. Yu, S. Wang, Forecasting corporate bankruptcy: Comparison of machine learning methods, *Expert Systems with Applications* 168 (2021) 114054. <https://doi.org/10.1016/j.eswa.2020.114054>.
- [2] H. Alaka, L. Oyedele, O. Akinade, M. Bilal, H. Owolabi, S. Ajayi, Machine learning in construction cost, time and risk analysis: A review, *Automation in Construction* 117 (2020) 103277. <https://doi.org/10.1016/j.autcon.2020.103277>
- [3] M. Qi, X. Yang, Deep learning for corporate default prediction, *Journal of Banking & Finance* 128 (2021) 106151. <https://doi.org/10.1016/j.jbankfin.2021.106151>
- [4] Y. Wang, R. Zhang, J. Li, Transformer-based corporate failure prediction using longitudinal financial data, *Decision Support Systems* 158 (2023) 114089. <https://doi.org/10.1016/j.dss.2023.114089>
- [5] X. Chen, Y. Liu, J. Sun, Imbalanced learning in credit risk modeling: A comparative study of resampling and cost-sensitive methods, *Applied Soft Computing* 121 (2022) 108765. <https://doi.org/10.1016/j.asoc.2022.108765>
- [6] G. Akbayrak, Y. Ozkan-Ozen, Y. Kazancoglu, Bankruptcy prediction using SMOTE and machine learning models, *Financial Innovation* 7 (2021) 58. <https://doi.org/10.1186/s40854-021-00298-y>
- [7] H. Li, Y. Feng, Z. Wang, SHAP-guided feature selection for ESG risk modeling in financial institutions, *Technological Forecasting and Social Change* 192 (2023) 122567. <https://doi.org/10.1016/j.techfore.2023.122567>
- [8] Y. Lu, D. Wu, Q. Hu, ESG risk forecasting using deep learning: A temporal attention model, *Finance Research Letters* 48 (2022) 103021. <https://doi.org/10.1016/j.frl.2022.103021>
- [9] R. Zhang, M. Chen, B. Liu, PanelGBM: Gradient boosting for panel data with fixed effects, *Knowledge-Based Systems* 260 (2023) 110589. <https://doi.org/10.1016/j.knosys.2023.110589>

- [10] J. Liu, K. Wang, X. Zhou, Time-aware undersampling for survival analysis in financial distress prediction, *IEEE Transactions on Neural Networks and Learning Systems* 35 (2024) 2105–2117. <https://doi.org/10.1109/TNNLS.2024.3356789>
- [11] D. Wu, Y. Lu, T. Zhang, Multimodal risk prediction using financial reports and textual signals: A hybrid deep learning approach, *Information Processing & Management* 60 (2023) 103456. <https://doi.org/10.1016/j.ipm.2023.103456>
- [12] Y. Feng, H. Li, X. Chen, Interpretable machine learning for corporate bankruptcy prediction: A SHAP-based analysis, *Decision Support Systems* 157 (2022) 113765. <https://doi.org/10.1016/j.dss.2022.113765>
- [13] N. Bussmann, P. Giudici, D. Marinelli, J. Papenbrock, Explainable AI in fintech risk management, *Frontiers in Artificial Intelligence* 3 (2020) 58. <https://doi.org/10.3389/frai.2020.00058>
- [14] L. Yu, W. Tang, C. Zhang, SHAP-based interpretation of bankruptcy models across industrial sectors, *International Journal of Forecasting* 38 (2022) 987–1001. <https://doi.org/10.1016/j.ijforecast.2021.08.005>
- [15] S. Kim, J. Park, Dynamic SHAP trajectories for longitudinal risk modeling in financial institutions, *Expert Systems with Applications* 236 (2024) 121890. <https://doi.org/10.1016/j.eswa.2024.121890>
- [16] S. M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, in: *Advances in Neural Information Processing Systems* 30 (2017) 4765–4774.
- [17] A. Fernandez, S. Garcia, M. Galar, R. C. Prati, B. Krawczyk, F. Herrera, *Learning from Imbalanced Data Sets*, Springer, 2018. <https://doi.org/10.1007/978-3-319-98074-4>
- [18] J. Xiao, Y. Wang, Z. Li, Ensemble methods for SME default prediction: A comparative study, *Financial Innovation* 8 (2022) 75. <https://doi.org/10.1186/s40854-022-00375-8>
- [19] D. Liang, C.-F. Tsai, Q. Wu, Credit risk assessment with LSTM and attention mechanism, *Expert Systems with Applications* 182 (2021) 115256. <https://doi.org/10.1016/j.eswa.2021.115256>
- [20] B. Ribeiro, M. dos Santos, Cost-sensitive learning for financial risk modeling in small and medium enterprises, *Applied Intelligence* 51 (2021) 5678–5693. <https://doi.org/10.1007/s10489-021-02567-z>
- [21] X. Chen, Y. Wang, Z. Liu, Integrating ESG into credit risk modeling: A machine learning approach, *Journal of Cleaner Production* 395 (2023) 136452. <https://doi.org/10.1016/j.jclepro.2023.136452>
- [22] Y. Wang, D. Wu, Y. Feng, Hybrid financial distress prediction using sentiment analysis and machine learning, *Knowledge-Based Systems* 256 (2022) 109123. <https://doi.org/10.1016/j.knosys.2022.109123>
- [23] Y. Lu, T. Zhang, Q. Hu, Temporal modeling of corporate risk with attention mechanisms: A deep learning framework, *Information Sciences* 648 (2023) 845–860. <https://doi.org/10.1016/j.ins.2023.08.045>
- [24] H. Zhang, M. Li, X. Chen, Interpretable deep learning for financial forecasting: A SHAP-enhanced approach, *Neurocomputing* 578 (2024) 127045. <https://doi.org/10.1016/j.neucom.2024.03.045>
- [25] I. Bayramli, V. Castro, Y. Barak-Corren, E. M. Madsen, M. K. Nock, J. W. Smoller, B. Y. Reis, Temporally informed random forests for suicide risk prediction, *Journal of the American Medical Informatics Association* 29 (2022) 62–71. <https://doi.org/10.1093/jamia/ocab225>
- [26] S. Vladov, V. Vysotska, V. Sokurenko, O. Muzychuk, M. Nazarkevych, V. Lytvyn, Neural Network System for Predicting Anomalous Data in Applied Sensor Systems, *Applied System Innovation* 7:5 (2024) 88. doi: 10.3390/asi7050088.