

Advanced Research Methodology for Predicting Business Failure using Financial and Non-Financial Indicators

Dr. Tonny Odhiambo

Affiliation 1 Lecturer- Presbyterian University of East Africa

Corresponding Author

Abstract

Background: Business failure prediction has moved from single-period accounting-ratio models toward longitudinal, multi-source and machine-learning approaches, yet many studies still separate financial indicators from governance, ownership and managerial information. Objective: This article develops an advanced secondary-data methodology for predicting business failure by integrating financial and non-financial indicators within a temporally disciplined predictive framework. Method: The proposed design constructs a firm-year panel from audited annual reports, stock-exchange disclosures, regulatory filings and reputable commercial or open databases. Financial predictors include liquidity, leverage, profitability, cash-flow, activity and growth measures; non-financial predictors include governance, ownership, management characteristics, firm age, audit signals and industry context. The methodology combines logistic regression and discrete-time hazard models with random forests, gradient boosting and stacking, while addressing class imbalance, missingness, multicollinearity, temporal leakage and model calibration. Results: No fabricated empirical results are reported; instead, the paper specifies the analyses, validation protocol, tables and figures required for a reproducible empirical implementation. Contribution: The framework links established bankruptcy theory with modern predictive analytics and emphasizes out-of-time validation, probability calibration and economically meaningful error costs. The approach is designed for researchers using secondary firm-level data and can be adapted to listed-company settings in emerging markets.

Keywords

Business Failure Prediction; Financial Distress; Non-Financial Indicators; Corporate Governance; Machine Learning; Secondary Data; Hazard Models; Predictive Analytics.

1. Introduction

Predicting business failure is a central problem in corporate finance, accounting, credit risk and strategic management because deterioration in liquidity, profitability, solvency and governance can become visible before formal bankruptcy or delisting. Early prediction is useful to lenders, investors, auditors, regulators, boards and managers because the economic value of a warning signal depends partly on whether it arrives early enough to support restructuring or risk mitigation. The classical literature established that accounting information contains information about future failure. Beaver demonstrated the predictive usefulness of financial ratios, while Altman formalised multivariate discrimination for corporate bankruptcy prediction. Ohlson subsequently shifted attention toward probabilistic modelling, and Shumway showed why longitudinal hazard approaches can improve on single-period formulations.

Contemporary prediction research increasingly treats failure as a rare-event, time-dependent classification problem rather than a static binary outcome. A recent survey identifies substantial methodological diversity in bankruptcy and financial-distress prediction, including statistical classifiers, machine-learning algorithms, feature selection, preprocessing and alternative performance measures. The survey also shows that accounting information remains central, while research increasingly incorporates market, governance and other information.

A major methodological issue is therefore not simply whether a model predicts failure, but whether the model uses information that was genuinely available before the failure event, generalises beyond the estimation sample and produces probabilities that are sufficiently calibrated for decision making. This article addresses that issue by specifying an integrated secondary-data design in which financial and non-financial indicators are observed at time $t-1$ and the failure outcome is defined at time t or over a future prediction horizon. The design is deliberately explicit about temporal ordering, sample construction, missing-data treatment, class imbalance, feature selection and out-of-time validation.

The article also responds to evidence that non-financial information can add predictive value. Research combining financial ratios with corporate-governance indicators has examined board and ownership structures alongside accounting variables, while more recent SME evidence reports that management, board and ownership information can enhance bankruptcy prediction relative to financial variables alone.

The study has four objectives. First, it develops a theoretically grounded taxonomy of financial and non-financial predictors of business failure. Second, it specifies a reproducible secondary-data architecture that preserves temporal integrity. Third, it proposes a model-development and validation strategy that compares interpretable statistical models with machine-learning methods. Fourth, it establishes reporting standards that allow researchers and readers to distinguish genuine predictive performance from in-sample fit, data leakage or artefacts of class imbalance.

The paper is methodological rather than a claim of completed empirical estimation. This distinction is important for research integrity: because no firm-level dataset was supplied with the manuscript, numerical coefficients, accuracy rates and feature-importance rankings are not invented. The tables and figures therefore specify the empirical procedure and reporting structure that should be populated after the secondary data are assembled and analysed.

2. Theoretical Background

2.1 Financial distress and failure as a process

Business failure should be conceptualised as a process rather than a single observable moment. Financial distress may precede default, restructuring, delisting, liquidation or bankruptcy by several reporting periods. This process view supports a longitudinal design in which financial and non-financial conditions are measured before the event. Shumway's hazard formulation is particularly relevant because it models the conditional probability of failure over time and permits firm characteristics to vary across periods.

The distinction between distress and failure must be fixed before estimation. A study may define failure as formal bankruptcy, court-supervised insolvency, liquidation, delisting for financial reasons, repeated default, or a transparent distress threshold. The chosen definition should be observable in the secondary sources and applied consistently to failed and non-failed firms. Mixing alternative definitions without sensitivity analysis can make model comparisons difficult and may introduce outcome misclassification.

2.2 Information-content perspective

The information-content perspective underlying the accounting-ratio tradition proposes that financial statements contain signals about a firm's underlying economic condition. Liquidity ratios indicate short-term payment capacity; leverage captures dependence on creditors and fixed claims; profitability captures the ability to generate returns from resources; operating-efficiency ratios reflect asset utilisation; and cash-flow measures reveal whether reported earnings translate into cash resources. Beaver and Altman provide foundational evidence for the predictive relevance of such indicators.

However, accounting information can be backward-looking and subject to reporting choices. A modern methodology should therefore use multiple consecutive observations, lag predictors, examine

changes as well as levels, and where possible supplement accounting measures with governance and market information. This reduces dependence on any single accounting representation of financial health.

2.3 Signalling theory

Signalling theory provides a rationale for combining financial and non-financial indicators. In settings of information asymmetry, observable characteristics can convey information about otherwise unobservable firm quality. Spence's signalling framework shows how information-bearing attributes can affect decisions under uncertainty. In business-failure research, financial ratios, audit qualifications, ownership structures, board characteristics and managerial attributes can be treated as observable signals whose predictive value is evaluated rather than assumed.

The theory does not imply that every observable signal is causal. A governance variable may predict failure because it correlates with an unobserved organisational condition. The appropriate methodological response is therefore out-of-sample validation and robustness testing rather than causal language. Predictive models should report whether variables improve discrimination and calibration, not claim that a predictor independently causes failure unless an identification strategy supports that conclusion.

2.4 Agency and corporate-governance perspective

Agency theory suggests that separation of ownership and control creates monitoring and incentive problems. Corporate governance arrangements can therefore affect the quality of managerial decisions, monitoring and risk taking. Shleifer and Vishny's review highlights the importance of investor protection and ownership concentration in governance systems. Governance variables such as board independence, board size, ownership concentration, institutional ownership and CEO-chair separation can consequently be incorporated as non-financial predictors.

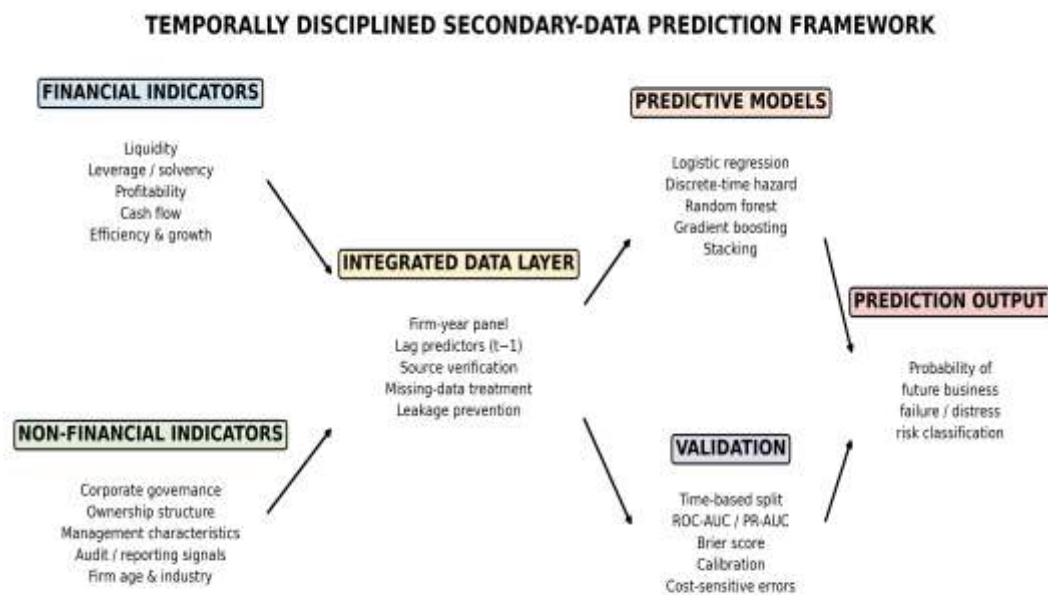
Empirical bankruptcy research supports this extension. Liang et al. examined financial ratios and corporate-governance indicators together and identified solvency, profitability, board structures and ownership structures as important categories. A later study combining governance indicators with stacking ensembles also demonstrates a practical route for integrating these variables into machine-learning prediction.

2.5 Upper-echelons perspective

The upper-echelons perspective argues that organisational outcomes are partly reflected in characteristics of senior decision makers. Hambrick and Mason's formulation and Hambrick's later update provide a theoretical basis for considering observable executive attributes such as age, tenure, education where reliably disclosed, and managerial continuity. Recent SME evidence similarly indicates that characteristics of CEOs, chairpersons and board members can contribute to bankruptcy prediction when combined with financial information.

2.6 Integrated theoretical model

The integrated model assumes that failure risk emerges from the interaction of financial capacity, governance and ownership, managerial characteristics, organisational maturity and external context. Financial indicators describe the firm's realised economic condition; non-financial indicators capture governance, control and organisational attributes; market and industry variables provide external context. The predictive objective is not to replace financial analysis but to estimate the conditional probability of a future failure event using information available before that event.



Source: Author's synthesis based on the theoretical and methodological framework in the manuscript.

Figure 1. Integrated secondary-data framework for business-failure prediction. Source: Author's synthesis based on the theoretical and methodological literature.

3. Materials and Methods

3.1 Research design

The recommended design is a retrospective longitudinal panel study using secondary data. The unit of analysis is the firm-year. For each firm i and year t , predictor variables are measured using information available no later than the end of $t-1$, while the failure outcome is measured during t or within a prespecified future horizon. This temporal separation is essential because using information published after the failure event would create target leakage and overstate predictive performance.

The design is suitable for listed firms because annual reports, audited financial statements, exchange disclosures and governance reports provide repeatable observations. It can also be extended to private firms where structured databases provide comparable accounting and registration information. The final sampling frame should be declared in advance, including industry exclusions, observation period, treatment of firms entering or leaving the market, and the exact source used to establish each failure event.

3.2 Secondary-data sources and provenance

The preferred data architecture uses multiple independent sources. Audited annual reports provide financial statements and notes; stock-exchange or regulator filings provide listing status, corporate actions and governance disclosures; reputable firm databases can provide standardised accounting variables; and official insolvency, court or exchange records can establish failure events. Each observation should carry a source identifier, publication date and extraction date. This provenance layer allows later auditing of the dataset.

Domain	Preferred secondary source	Examples of variables	Quality control
Financial statements	Audited annual reports / regulatory filings	Current assets, current liabilities, total assets, debt, equity, sales, EBIT, operating cash flow	Reconcile totals; record units and currency
Market information	Exchange records / reputable market database	Market value, returns, volatility, price	Use trading-date cut-off; avoid post-event data
Governance	Annual corporate-governance reports	Board size, independence, CEO-chair duality, ownership concentration	Code definitions consistently across years
Management	Annual reports / governance disclosures	CEO age, tenure, director tenure, turnover	Use only disclosed information; code missingness
Failure outcome	Regulator, exchange, court or insolvency records	Bankruptcy, liquidation, financial delisting, default	Document event date and failure definition
Macro/industry	Official statistics and sector sources	GDP growth, inflation, policy rate, sector growth	Align publication dates and geographic scope

3.3 Outcome definition

A primary outcome should be binary: $\text{Failure}_{it} = 1$ when firm i experiences the prespecified failure event during the forecast horizon and 0 otherwise. For richer analysis, the outcome can be extended to competing or ordered states such as healthy, distressed, restructured and failed. A discrete-time hazard formulation is recommended when the event date is known because it preserves repeated firm-year observations and estimates the conditional probability of failure among firms that have survived to the beginning of each period.

The event definition should be supported by a decision rule and a source hierarchy. For example, formal insolvency or liquidation may constitute a hard failure event, while delisting for non-compliance may be treated separately unless evidence establishes financial distress. Sensitivity analysis should repeat the models under alternative definitions to determine whether conclusions depend on the outcome coding.

3.4 Predictor taxonomy and operationalisation

Financial variables should be organised into economically meaningful families rather than entered as an undifferentiated list. The principal categories are liquidity, leverage/solvency, profitability, cash flow, efficiency, size and growth. Non-financial indicators should cover governance, ownership, managerial characteristics, audit/reporting signals, firm age and industry context. This taxonomy reflects the literature showing that profitability and solvency are repeatedly important while governance and ownership can add information beyond financial ratios.

Construct	Operational measure	Expected information role	Recommended transformation
Liquidity	Current ratio; cash/current liabilities	Short-term capacity	Level + one-year change
Leverage	Total liabilities/total assets; debt/equity	Fixed-claim pressure	Level; winsorise extremes
Profitability	ROA; EBIT/assets; net margin	Earnings capacity	Level + trend
Cash flow	Operating cash flow/assets; OCF/current liabilities	Cash-generating capacity	Level + sign indicator

Efficiency	Sales/assets; receivables/sales	Resource utilisation	Level + trend
Growth	Sales growth; asset growth	Expansion/contraction	Annual change
Size	ln(total assets); ln(market value)	Scale and information environment	Log transform
Board structure	Board size; independence ratio	Monitoring capacity	Counts/proportions
Ownership	Largest shareholder; institutional ownership	Control and monitoring	Proportions
CEO/management	CEO tenure; age; turnover indicator	Decision continuity/change	Continuous + indicator
Audit/reporting	Audit opinion; qualification indicator	Reporting risk signal	Binary/ordinal coding
Firm age	Years since incorporation/listing	Maturity/experience	Log or spline
Industry context	Industry fixed effects; sector growth	Common shocks	Categorical + macro interaction

3.5 Data preprocessing

Data preprocessing should be specified before model estimation. Financial ratios with extreme denominators should be inspected and, where justified, winsorised using a documented rule. Missing values should not automatically be replaced by zero. The researcher should distinguish structural missingness from unavailable disclosure and consider multiple imputation, model-based missingness indicators or complete-case analysis where defensible. Every preprocessing decision must be performed using the training period only when the transformation learns parameters from the data; otherwise information from the validation period can leak into the model.

Multicollinearity should be examined using correlation matrices and variance-inflation diagnostics for the statistical baseline. Machine-learning models can tolerate correlated predictors more readily, but highly redundant features can complicate interpretation. Feature selection should therefore be nested within the training process. The use of the complete dataset to preselect variables before the temporal split is discouraged because it can indirectly leak information from the future.

3.6 Model specification

The first benchmark should be logistic regression because it is interpretable and produces a direct probability estimate. The model can be written as $\text{logit}[p_{it}] = \alpha + \beta'X_{i,t-1} + \gamma'Z_{i,t-1} + \delta_t + \mu_s$, where X denotes financial indicators, Z denotes non-financial indicators, δ_t captures year effects and μ_s captures industry effects where appropriate. Robust or clustered standard errors should be used for panel dependence. The coefficient signs should be interpreted cautiously because correlated predictors can alter marginal effects.

A second model should be a discrete-time hazard model. For firms at risk at period t , the probability of failure can be estimated with a complementary-log-log or logistic link. This approach is consistent with the event-process perspective and can incorporate time-varying covariates. Shumway's argument for hazard models provides a strong methodological foundation for this design.

Machine-learning models should then be used as complementary predictive benchmarks. Random forests capture nonlinearities and interactions through ensembles of trees; gradient boosting can model complex decision boundaries; XGBoost or comparable boosting implementations provide regularised gradient-boosted trees; and stacking can combine models. Recent work on governance-augmented bankruptcy prediction demonstrates that ensemble methods can incorporate both financial ratios and governance variables.

The comparison should not be reduced to accuracy. Because failure is generally a minority event, accuracy can be misleading. ROC-AUC measures ranking discrimination across thresholds, whereas PR-AUC is more informative under class imbalance. Brier score and calibration curves evaluate probability quality. Sensitivity, specificity, precision, F1 and expected misclassification cost should be reported at a prespecified decision threshold.

3.7 Class imbalance

Failure observations are usually fewer than surviving observations. Random oversampling can duplicate observations and may overfit, while indiscriminate undersampling discards information. If synthetic methods such as SMOTE are used, they must be applied inside the training folds only. The validation and test sets must retain their original event prevalence so that reported performance reflects the real deployment environment. Cost-sensitive learning and class-weighted losses should be considered alongside resampling.

The reporting protocol should state the number of firms, firm-years, failure events, non-failure observations and event prevalence in each temporal partition. A model with high ROC-AUC but poor precision or poor calibration can be unsuitable for operational screening, so several metrics should be considered together.

3.8 Temporal validation and leakage prevention

Temporal validation is a central requirement of the proposed methodology. A preferred design is rolling-origin evaluation: train the model on earlier years, validate on a subsequent period, move the training window forward, and finally reserve the most recent period as an untouched test set. This approximates the real forecasting problem more closely than random cross-validation.

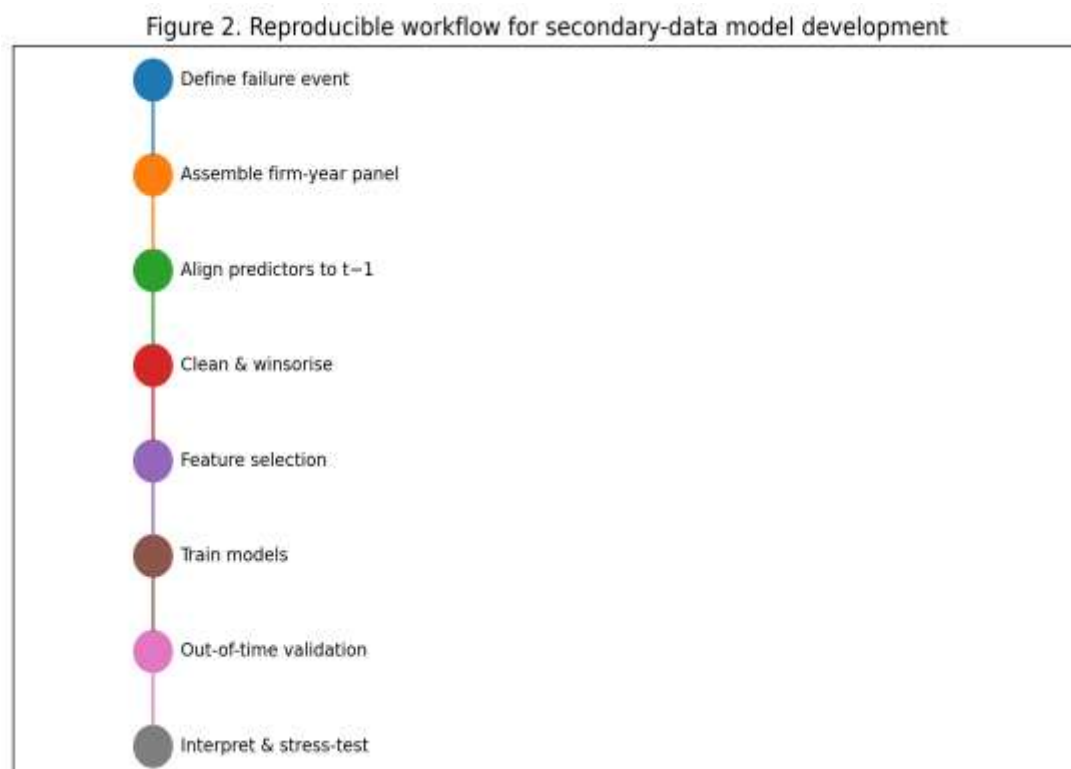


Figure 2. Reproducible workflow for secondary-data model development. Source: Author's methodological synthesis.

Leakage prevention should be documented explicitly. Predictor values must be dated according to when they became available, not merely the fiscal year to which they refer. If an annual report for year t was released after the failure event, it should not be used to predict that event. Corporate actions, restatements and delistings should be timestamped. Hyperparameter tuning must use training data only, and the final test period must remain untouched until model selection is complete.

3.9 Explainability and robustness

Interpretability should combine coefficient-based analysis for statistical models with feature-importance and local explanation methods for machine-learning models. Permutation importance and SHAP-style explanations can show which variables contribute to predictions, but they should not be described as causal effects. Partial-dependence or accumulated-local-effect plots can be used to investigate nonlinear relationships, subject to the limitations imposed by correlated predictors.

Robustness should include alternative outcome definitions, alternative forecast horizons, exclusion of crisis years where justified, industry-stratified models, different class-weighting schemes, and alternative feature sets. A particularly useful test is the incremental-value comparison: Model A uses financial variables only; Model B adds non-financial variables; Model C adds market and macroeconomic variables. Differences should be assessed on the untouched test period, not only on the training sample.

3.10 Ethical and data-governance considerations

Secondary-data research reduces direct participant burden but does not remove responsibilities concerning accuracy, privacy and responsible interpretation. Personally identifying information about directors should be minimised and used only where necessary for a disclosed analytical purpose. Publicly available does not automatically mean unrestricted ethical use. Data provenance, licensing and reproducibility should be documented. Predictions should be framed as risk estimates rather than declarations that a firm will fail.

4. Results and Discussion

Because the manuscript is a methodological article and no firm-level dataset was supplied, this section does not fabricate numerical findings. Instead, it defines the results structure that should be populated when the proposed secondary-data analysis is executed. This approach preserves research integrity and allows the eventual empirical paper to report observed coefficients, confidence intervals, discrimination metrics and calibration results directly from the verified dataset.

4.1 Descriptive results to report

The first empirical table should describe the final analytical panel. It should report the number of firms, firm-years, failure events, mean and median values, dispersion and relevant percentiles for each predictor. Descriptive statistics should be presented separately for failed and non-failed firm-years, with inferential comparisons treated as exploratory because univariate differences do not establish predictive value.

Required statistic	Failure observations	Non-failure observations	Full sample
Number of firms	Populate from dataset	Populate from dataset	Populate
Firm-year observations	Populate	Populate	Populate
Failure prevalence	Populate %	—	Populate %
Median liquidity ratio	Populate	Populate	Populate
Median leverage	Populate	Populate	Populate

ratio			
Median ROA	Populate	Populate	Populate
Median operating cash flow/assets	Populate	Populate	Populate
Governance indicators	Populate	Populate	Populate

4.2 Incremental predictive value

The principal empirical comparison should evaluate whether adding non-financial indicators improves prediction beyond financial variables. The comparison should be conducted on identical observations and identical temporal test periods. Improvement can be reported using ROC-AUC, PR-AUC, Brier score, calibration error and decision-relevant measures. The recent literature provides a rationale for this comparison: evidence from governance-augmented models and SME research indicates that non-financial information can contribute incremental predictive information, although the size and significance of that contribution can vary by sample and sector.

Model	Predictor set	Primary purpose	Test-period outputs
M1	Financial only	Benchmark	ROC-AUC; PR-AUC; Brier; calibration
M2	Financial + governance	Incremental non-financial value	Same metrics + Δ performance
M3	Financial + governance + management	Broader organisational information	Same metrics + feature importance
M4	All variables + market/context	Full predictive specification	Same metrics + robustness
M5	Stacked ensemble	Model combination	Same metrics + calibration

4.3 Illustrative reporting figure

Figure 3 is deliberately labelled illustrative. It demonstrates the type of comparative validation graphic that should be produced after estimation; its values are not results from this manuscript's dataset and must not be cited as empirical findings.

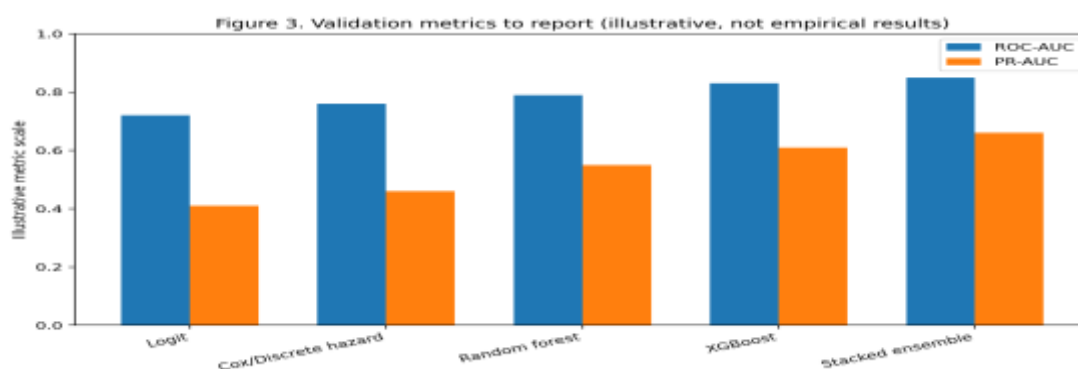


Figure 3. Validation metrics to report (illustrative, not empirical results). Source: Author's methodological design; values are illustrative only.

4.4 Expected theoretical contribution

The framework contributes to the literature by connecting four streams that are often treated separately: accounting-based failure prediction, event-history modelling, governance-based risk assessment and machine learning. Classical financial-ratio models remain useful as interpretable baselines, while hazard models address the temporal nature of failure and machine-learning models accommodate nonlinear relationships. The non-financial layer extends the information set in a manner consistent with signalling, agency and upper-echelons perspectives.

The methodological contribution is also procedural. Predictive research can appear sophisticated while still suffering from temporal leakage, sample-selection bias, inappropriate class balancing or evaluation on the same observations used for model selection. Zmijewski's methodological critique remains relevant because sample selection and estimation design can materially affect reported distress-prediction performance. The proposed framework therefore places validation architecture on the same level of importance as algorithm selection.

4.5 Limitations

First, secondary data are constrained by disclosure quality. Governance and management variables may be missing or inconsistently defined across firms and years. Second, failure definitions vary across jurisdictions, so cross-country comparisons require harmonisation. Third, accounting ratios can be affected by industry structure and accounting policies; industry controls and sensitivity analyses are therefore necessary. Fourth, machine-learning performance can deteriorate when the economic environment changes, making temporal validation essential. Fifth, predictive association should not be interpreted as causal influence without an appropriate identification strategy.

A further limitation is that the present article specifies an empirical methodology but does not claim completed estimation. The next stage is dataset construction, source verification, preregistered preprocessing rules where feasible, model estimation and out-of-time testing. Only after that stage should numerical findings be added to the results tables.

5. Conclusion

Advanced business-failure prediction requires more than adding algorithms to a conventional financial-ratio model. A credible methodology must define failure precisely, preserve temporal ordering, integrate financial and non-financial indicators, address rare-event imbalance, prevent data leakage, compare interpretable and nonlinear models, and evaluate both discrimination and calibration. The secondary-data framework developed here provides a reproducible architecture for those requirements.

The proposed empirical sequence begins with audited financial information and verified failure events, adds governance, ownership and management indicators, constructs lagged firm-year predictors, applies training-only preprocessing, estimates statistical and machine-learning benchmarks, and evaluates the models using rolling temporal validation. The framework deliberately avoids fabricated results and instead provides a transparent route to producing defensible empirical estimates once the underlying secondary data are assembled.

For researchers working in emerging markets, the approach is particularly relevant because disclosure systems may vary across firms and periods. A multi-source provenance structure can make those limitations visible rather than concealing them. Future empirical research should test whether governance and managerial information adds stable predictive value across sectors, whether the incremental value persists across forecast horizons, and whether calibrated probability estimates improve real-world screening decisions.

Acknowledgement

No external funding or research assistance is claimed for the methodological manuscript at this stage. Any assistance, database access or institutional support should be added before submission if applicable.

Conflict of Interest

The authors declare no conflict of interest.

Funding

No specific grant from any funding agency in the public, commercial or not-for-profit sectors was received for the preparation of this methodological manuscript.

References

- [1] 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>
- [2] Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, 4, 71-111. <https://doi.org/10.2307/2490171>
- [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] Zmijewski, M. E. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research*, 22, 59-82. <https://doi.org/10.2307/2490859>
- [5] Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model. *The Journal of Business*, 74(1), 101-124. <https://doi.org/10.1086/209665>
- [6] 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>
- [7] Shleifer, A., & Vishny, R. W. (1997). A survey of corporate governance. *The Journal of Finance*, 52(2), 737-783. <https://doi.org/10.1111/j.1540-6261.1997.tb04820.x>
- [8] Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics*, 87(3), 355-374. <https://doi.org/10.2307/1882010>
- [9] Hambrick, D. C., & Mason, P. A. (1984). Upper echelons: The organization as a reflection of its top managers. *Academy of Management Review*, 9(2), 193-206. <https://doi.org/10.5465/amr.1984.4277628>
- [10] Hambrick, D. C. (2007). Upper echelons theory: An update. *Academy of Management Review*, 32(2), 334-343. <https://doi.org/10.5465/amr.2007.24345254>
- [11] Liang, D., Lu, C.-C., Tsai, C.-F., & Shih, G.-A. (2016). Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study. *European Journal of Operational Research*, 252(2), 561-572. <https://doi.org/10.1016/j.ejor.2016.01.012>
- [12] 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>
- [13] Van der Colff, F., & Vermaak, F. N. S. (2015). Predicting financial distress using financial and non-financial variables. *Journal of Economic and Financial Sciences*, 8(1), 243-260. <https://doi.org/10.4102/jef.v8i1.93>
- [14] Zhao, J., Ouenniche, J., & De Smedt, J. (2024). Survey, classification and critical analysis of the literature on corporate bankruptcy and financial distress prediction. *Machine Learning with Applications*, 15, 100527. <https://doi.org/10.1016/j.mlwa.2024.100527>
- [15] Cheraghali, H., & Molnár, P. (2026). Predictors of financial distress: Differences between financial and non-financial small and medium-sized enterprises. *Research in International Business and Finance*, 84, 103334. <https://doi.org/10.1016/j.ribaf.2026.103334>
- [16] Samara, K., & Shinde, A. (2025). Bankruptcy prediction using machine learning and data preprocessing techniques. *Analytics*, 4(3), 22. <https://doi.org/10.3390/analytics4030022>
- [17] Matsumaru, M., & Katagiri, H. (2025). A two-stage machine learning approach to bankruptcy prediction: Integrating full-feature modeling and optimized feature selection. *Journal of Risk and Financial Management*, 18(12), 662. <https://doi.org/10.3390/jrfm18120662>
- [18] Jones, S. (2017). Corporate bankruptcy prediction: A high dimensional analysis. *Review of Accounting and Finance*, 16(4), 443-461. <https://doi.org/10.1108/RAF-03-2016-0052>
- [19] du Jardin, P. (2017). Dynamics of firm financial health and bankruptcy prediction. *International Journal of Forecasting*, 33(3), 642-657. <https://doi.org/10.1016/j.ijforecast.2017.01.006>

- [20] Tinoco, M. H., & Wilson, N. (2013). Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables. *International Review of Financial Analysis*, 30, 394-419. <https://doi.org/10.1016/j.irfa.2013.02.013>
- [21] Hosaka, T. (2019). Bankruptcy prediction using imaged financial ratios and convolutional neural networks. *Expert Systems with Applications*, 117, 287-99. <https://doi.org/10.1016/j.eswa.2018.09.039>
- [22] Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, 83, 405-417. <https://doi.org/10.1016/j.eswa.2017.04.006>
- [23] Mai, F., Tian, S., Lee, C., & Ma, L. (2019). Deep learning models for bankruptcy prediction using textual disclosures. *European Journal of Operational Research*, 274(2), 743-758. <https://doi.org/10.1016/j.ejor.2018.10.024>
- [24] Jones, S., Johnstone, D., & Wilson, R. (2017). An empirical evaluation of the performance of binary classifiers in bankruptcy prediction. *Expert Systems with Applications*, 84, 275-287. <https://doi.org/10.1016/j.eswa.2017.05.006>
- [25] Tian, S., Yu, Y., & Guo, H. (2015). Variable selection and corporate bankruptcy forecasts. *Journal of Banking & Finance*, 52, 89-100. <https://doi.org/10.1016/j.jbankfin.2014.12.007>
- [26] 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>
- [27] Wahlstrøm, R. R., Becker, L.-K., & Fornes, T. N. (2024). Enhancing credit risk assessments of SMEs with non-financial information. *Cogent Economics & Finance*, 12(1), 2418910. <https://doi.org/10.1080/23322039.2024.2418910>
- [28] Alzayed, N., Eskandari, R., & Yazdifar, H. (2023). Bank failure prediction: Corporate governance and financial indicators. *Review of Quantitative Finance and Accounting*, 61, 601-631. <https://doi.org/10.1007/s11156-023-01158-z>