Multi-Modal Financial Distress Prediction Through the Fusion of Price Dynamics, Macroeconomic Signals, and Corporate Disclosures

Navin Venkataraman
Department of Computer Science, University of Central Florida, Orlando, FL, USA.
venkataraman923@ucf.edu

Bennett Kerry
Department of Computer Science, University of North Texas, Denton, TX, USA.
terrybennett@unt.edu

Abstract

Financial distress prediction remains a central challenge in quantitative finance, risk management, and regulatory oversight. Traditional models relying on a single data modality, such as accounting ratios or market prices, suffer from limited predictive power, temporal instability, and an inability to capture the multi-faceted nature of economic crises. This paper proposes a multi-modal framework that fuses three distinct signal families: high-frequency price dynamics, macroeconomic indicators, and structured corporate disclosures. We argue that no single information channel is sufficient to anticipate the complex interplay of firm-specific vulnerabilities, systemic shocks, and informational frictions. The proposed architecture integrates temporal attention mechanisms and cross-modal fusion layers to learn hierarchical representations that reflect both micro-level financial health and macro-level stress propagation. We discuss the structural trade-offs inherent in model design, including the tension between predictive accuracy and interpretability, the risks of data leakage in benchmark construction, and the challenges of deploying such systems across heterogeneous regulatory environments. Special attention is given to the evaluation of market-stress early warning systems, where traditional backtesting often conflates statistical fit with economic credibility. Drawing on recent advances in leakage-safe benchmark design and attention-enhanced reinforcement learning, we outline a governance framework that promotes robustness, fairness, and sustainability. The paper concludes with policy implications for systemic risk monitoring and the responsible integration of multi-modal artificial intelligence into financial infrastructure. Our analysis underscores that the fusion of diverse signals, while technically demanding, offers the most promising path toward resilient distress prediction systems that can operate under extreme uncertainty.

Keywords

financial distress prediction, multi-modal fusion, price dynamics, macroeconomic signals, corporate disclosures, systemic risk, data leakage, benchmark design, artificial intelligence governance.

1. Introduction

The accurate prediction of financial distress has long been a cornerstone of credit risk assessment, portfolio management, and macroprudential regulation. From the seminal work of Altman [1] using discriminant analysis on accounting ratios to the logistic regression models of Ohlson [2], the field has evolved primarily through the refinement of single-modality inputs. However, the financial crises of the past two decades, including the 2008 global financial crisis and the COVID-19 induced market disruptions, have exposed fundamental limitations in models that rely exclusively on balance sheet data or market prices. These events demonstrated that distress often propagates through interconnected channels, where a sudden deterioration in macroeconomic conditions, a sharp decline in equity prices, and a loss of confidence in corporate disclosures can reinforce each other in nonlinear ways. A multi-modal approach that simultaneously processes price dynamics, macroeconomic signals, and corporate textual disclosures is therefore not merely an incremental improvement but a structural necessity.

The primary contribution of this paper is to propose a system architecture that fuses these three heterogeneous data modalities into a unified predictive framework. We adopt a system-level perspective, emphasizing the design choices, trade-offs, and deployment challenges that arise when integrating high-frequency market data, low-frequency macroeconomic releases, and irregularly published corporate reports. Each modality presents distinct temporal granularities, noise characteristics, and informational biases. Price dynamics, for instance, reflect immediate market sentiment and arbitrage activity but are contaminated by microstructure noise and speculative frictions. Macroeconomic signals, such as GDP growth, unemployment rates, and yield curve slopes, capture slow-moving regime shifts but suffer from publication lags and revisions. Corporate disclosures, including annual reports and management discussion sections, offer forward-looking textual evidence of financial constraints and strategic pivots, yet they are subject to impression management and selective disclosure.

More importantly, the fusion of these modalities introduces critical issues related to data leakage, benchmark credibility, and fairness. Recent work by the author of [16] has demonstrated that standard benchmark designs for market-stress early warning are often plagued by look-ahead bias and temporal dependencies that inflate reported performance. Similarly, the author of [3] proposed a PCA-APT stress index that attempts to disentangle systematic risk factors from idiosyncratic distress signals, but the integration of such indices into multi-modal models requires careful alignment of time horizons. Our analysis builds on these foundations by advocating for leakage-aware evaluation protocols and economically meaningful loss functions. We also draw on the insights of [4], who compared transformer-based and classical models for financial time-series forecasting, concluding that attention mechanisms offer superior ability to capture long-range dependencies but at the cost of increased computational overhead and opacity.

The structure of this paper is as follows. Section 2 reviews the relevant literature across financial distress prediction, multi-modal learning, and the economics of corporate disclosure. Section 3 describes the three data modalities and the architectural decisions involved in their fusion. Section 4 discusses the fusion mechanisms, including temporal attention and cross-modal alignment. Section 5 addresses evaluation strategy, with a focus on benchmark integrity. Section 6 examines governance, fairness, and deployment considerations. Section 7 outlines future research directions and policy implications. Section 8 concludes.

2. Background and Related Work

Financial distress prediction has been studied extensively through the lens of accounting-based models, market-based models, and more recently, machine learning approaches. Altman's Z-score [1] established a discriminant function using working capital, retained earnings, earnings before interest and taxes, market value of equity, and sales. Ohlson's O-score [2] extended this to logistic regression using financial ratios and size variables. These models remain widely used in practice but exhibit significant degradation in predictive performance during periods of structural change, such as recessions or regulatory reforms. More sophisticated models incorporate market variables, such as stock return volatility, distance to default, and credit spreads, as in the Merton-based approach of [5].

The introduction of machine learning has dramatically expanded the feature space. Support vector machines, random forests, and gradient boosting methods have been applied to distress prediction with reported improvements in accuracy [6]. However, these models typically rely on a single modality, often a static set of accounting ratios or a fixed window of price returns. The temporal aspect is usually addressed through rolling windows or time-series aggregation, but the integration of multiple signal types remains underdeveloped. Recent advances in deep learning, particularly recurrent neural networks and transformers, have enabled the modeling of sequential dependencies in financial data [7]. Yet the challenge of fusing heterogeneous modalities with different sampling frequencies and noise levels persists.

From the macroeconomic perspective, distress prediction has been linked to early warning systems for systemic risk. The literature on macroprudential indicators, such as the credit-to-GDP gap and asset price misalignments, provides a complementary layer of information [8]. The author of [9] introduced a PCA-APT stress index that aggregates principal components of asset pricing factors to measure latent macroeconomic stress. This index, when aligned with firm-level data, offers a systematic way to incorporate risk premia that vary with the business cycle. However, the integration of such indices into micro-level prediction models requires careful consideration of scale, frequency, and causal direction.

Corporate disclosures, particularly the narrative sections of annual reports, have been mined for sentiment, tone, and forward-looking statements. Text mining techniques, including bag-of-words, latent Dirichlet allocation, and transformer-based language models, have been used to extract signals of financial constraint, earnings manipulation, and managerial optimism [10]. The author of [11] examined the impact of financial constraints on ESG ratings using Chinese stock market data, highlighting how disclosure quality itself can be a marker of distress. The fusion of textual data with quantitative signals is non-trivial because text is often produced at lower frequency and with intentional ambiguity.

Multi-modal learning in finance is an emerging field. Existing work typically concatenates features from different sources or uses separate encoders that are combined at a late stage [12]. Attention-based fusion mechanisms, which allow the model to dynamically weigh the importance of each modality over time, have shown promise in other domains such as medical diagnosis and autonomous driving. Their application to financial distress prediction, however, must account for the unique properties of financial data, including non-stationarity, heavy tails, and adversarial manipulation. The author of [13] explored attention-enhanced reinforcement learning for dynamic portfolio optimization, demonstrating that cross-modal attention can improve out-of-sample performance. This idea can be adapted to distress prediction by allowing the model to selectively attend to price jumps, macroeconomic shocks, or disclosure changes depending on the prevailing regime.

3. System Architecture and Data Modalities

The proposed system is designed to operate as an end-to-end pipeline that ingests three distinct data streams, transforms them into a shared representational space, and produces a time-varying distress probability for each firm. The architecture must handle asynchronous updates: price data arrives at sub-second frequencies, macroeconomic indicators are released monthly or quarterly with revisions, and corporate disclosures appear annually or semi-annually, often with delays. The first architectural decision is the synchronization of these signals. We adopt a fixed-time-step framework with daily granularity, where low-frequency signals are forward-filled and high-frequency signals are aggregated to daily returns and volatility measures. This choice introduces a trade-off between temporal resolution and signal distortion. Daily aggregation loses intraday contagion effects but reduces noise and computational cost. More sophisticated approaches could use event-driven architectures, but for system-level deployment, regular time steps simplify monitoring and backtesting.

Price dynamics are represented by a set of derived features including daily log returns, realized volatility estimated from five-minute intraday data, relative trading volume, and the firm's beta relative to a market index. These features capture market-implied distress through sharp price declines, increased uncertainty, and heightened trading activity. However, price dynamics alone are prone to false signals during flash crashes or liquidity events that are not fundamentally driven. To mitigate this, the system also incorporates a market-wide stress index derived from macro factor models, as proposed by the author of [14]. This index serves as a contextual variable that scales firm-level signals by the prevailing level of systematic stress.

Macroeconomic signals are collected from a curated set of indicators including the yield curve slope, corporate bond spreads, unemployment claims, industrial production growth, and consumer confidence indices. These variables are themselves the output of complex estimation processes and come with confidence intervals and revision histories. A robust system must account for the fact that the most recent macro data point may be a preliminary estimate with significant uncertainty. We therefore include an uncertainty mask that down-weights recent macro signals during training and inference. The macro modality is encoded using a temporal convolutional network that captures the multi-scale nature of business cycles, as opposed to recurrent architectures that may amplify noise.

Corporate disclosures are processed using a pre-trained transformer language model fine-tuned on a corpus of financial reports. Key features extracted include the proportion of uncertain and negative words, the frequency of forward-looking statements, and topic proportions related to leverage, liquidity, and litigation. The textual modality also includes the length of the management discussion section and the number of risk factors listed. A critical challenge is the irregular issuance of reports; some firms may delay filings when under financial stress, creating missing data that is not random. The system handles missing disclosures by using a learnable placeholder embedding that is conditioned on the firm's recent price and macro context, following the approach of [15].

4. Fusion Mechanisms and Model Design

The core of the proposed system is a cross-modal fusion layer that integrates the encoded representations from each modality at each time step. A naive concatenation of features would ignore the temporal dependencies and cross-modal interactions that are essential for capturing the propagation of distress. Instead, we employ a multi-head attention mechanism that allows each modality to attend to the others at every time step, producing a context-aware representation. Specifically, the price representation attends to macro and text representations to adjust for systematic versus idiosyncratic movements; the macro representation attends to price and text to detect regime change signals; the text representation attends to price and macro to contextualize the disclosure tone.

This attention-based fusion introduces a computational overhead that scales quadratically with the number of time steps and modalities. For a system deployed across thousands of firms, this may become prohibitive. A structural trade-off exists between the richness of cross-modal interactions and the latency of inference. For real-time early warning systems, we recommend a staggered fusion schedule: full attention during daily batch re-estimation, but a light-weight linear projection during intraday updates. Such a design balances responsiveness with fidelity.

The final distress probability is produced by a feed-forward classifier that takes the fused representation as input. The classifier is trained using a binary cross-entropy loss with a time-weighted penalty that emphasizes recent observations. However, financial distress is a rare event, and class imbalance can lead to overconfident models that predict low distress probabilities for most firms. To address this, we incorporate a cost-sensitive learning scheme and a threshold selection procedure based on the precision-recall trade-off. Additionally, we include a regularization term that penalizes large changes in predicted probabilities between consecutive time steps, encouraging temporal stability.

A central concern in multi-modal fusion is the risk that one modality may dominate the learning process. For example, price data, with its high frequency and low acquisition cost, may cause the model to ignore textual signals that are more informative for long-term distress. We mitigate this through modality dropout during training, where each modality is randomly masked with a certain probability, forcing the model to rely on the remaining signals. This technique, adapted from [17], improves robustness and prevents over-reliance on any single information channel.

5. Evaluation Strategy and Benchmark Integrity

Evaluating a multi-modal distress prediction system is fraught with pitfalls. Standard backtesting procedures that split data chronologically into training and test sets are vulnerable to temporal data leakage, where future information contaminates past predictions through look-ahead in feature construction. For instance, using the full history of stock prices to compute a moving average that then serves as a feature for a date in the middle of the sample introduces information from later returns. The author of [16] proposed a leakage-safe benchmark design that explicitly defines the point of information availability for each feature and strictly enforces that no future data is used. This design is particularly important for multi-modal systems because macroeconomic revisions and corporate filings often have delayed publication dates. A feature that uses the final revised GDP figure for a quarter may not be available at the time of prediction if the first estimate was released later. Therefore, we advocate for a rigorous timestamp alignment that records the earliest possible date each data point could have been known to a market participant.

Beyond leakage, the economic credibility of evaluation metrics must be considered. A model that achieves high area under the receiver operating characteristic curve may still fail to identify distress early enough for preventive action. The author of [14] introduced a PCA-APT stress index that is designed to be evaluated under economically motivated loss functions, such as the cumulative loss from missed defaults versus false alarms. We adopt a similar approach by measuring system performance using a cost-based metric that assigns higher penalties to late predictions and false positives that trigger unnecessary intervention. This aligns model optimization with the true objectives of regulators and financial institutions.

Another concern is the representativeness of the test set. Most financial datasets cover periods of relative stability, with only a few crisis episodes. A model trained on such data may overfit to the idiosyncrasies of a single crisis and fail to generalize to new types of shocks. We recommend stress testing using synthetic scenarios generated from counterfactual macroeconomic simulations. This approach, while complex, provides a more robust assessment of out-of-distribution performance. The author of [4] demonstrated that transformer-based models, while expressive, are particularly sensitive to distributional shift, and careful regularization is needed.

6. Governance, Fairness, and Deployment Considerations

The deployment of a multi-modal distress prediction system in real-world financial infrastructure raises profound governance questions. First, the opacity of deep learning models creates a tension with the regulatory requirement for explainability. Financial institutions and central banks must be able to justify the predictions they use for capital allocation, stress testing, or supervisory actions. Our system incorporates a post-hoc explanation module that uses attention weights and integrated gradients to attribute each prediction to specific features across modalities. While not fully causal, these explanations provide a layer of accountability.

Fairness is another critical dimension. If the model systematically underestimates distress for firms that are large or politically connected, it may perpetuate existing inequalities and amplify systemic risk. Bias can arise from imbalanced training data, where certain industries or regions are underrepresented. For example, the author of [11] found that financial constraints have a heterogeneous impact on ESG ratings across Chinese stock market segments, and a model trained on a broad sample might not capture these nuances. We advocate for fairness audits that compare false negative rates across subgroups defined by firm size, industry, and geographic location. If disparities are found, the system should be reweighted or augmented with additional features that correct for the bias.

Sustainability of the system over time is a further concern. Financial markets evolve, new types of disclosures emerge, and macroeconomic regimes shift. A static model trained once will rapidly become obsolete. We therefore propose a lifecycle management framework that includes periodic retraining, drift detection, and automated rollback procedures. The system should also be designed to operate under adverse conditions, such as market closures, data feed failures, or cyberattacks. Redundant data sources and fallback heuristics should be embedded to ensure continuity.

Policy implications extend to systemic risk monitoring. A multi-modal early warning system could be used by regulatory authorities to identify firms that are likely to default under various stress scenarios. However, if such a system becomes widely adopted, it may create feedback loops where predictions themselves alter market behavior. For instance, a publicized distress prediction could trigger a run on a firm’s stock, accelerating its failure. This reflexivity must be accounted for in the governance framework. One approach is to keep the system internal to the regulator and not disclose individual firm scores, but rather use them to inform aggregate risk assessments.

7. Future Directions and Policy Implications

The fusion of price dynamics, macroeconomic signals, and corporate disclosures opens several avenues for future research. One promising direction is the incorporation of alternative data sources, such as news sentiment, social media activity, and supply chain connections. These modalities add further temporal and semantic complexity but also increase the risk of noise and manipulation. The author of [13] showed that attention-enhanced reinforcement learning can dynamically adjust portfolio weights based on cross-modal signals; a similar framework could be used to adapt the fusion weights in real-time, making the system more responsive to changing market conditions.

Another research frontier is the development of causal models that go beyond correlation. Current fusion techniques learn statistical associations, but distress causation often involves feedback loops and latent confounders. For example, a decline in price may be caused by a macroeconomic shock, but the price decline itself may worsen the firm's credit rating, which then appears in disclosures. Causal discovery algorithms that operate on mixed-frequency time series could disentangle these pathways, leading to more robust predictions and better policy interventions.

Policy implications are substantial. Regulators could mandate that financial institutions incorporate multi-modal signals into their internal capital adequacy assessments. However, this would require standardization of data formats, disclosure timeliness, and model validation procedures. International coordination is needed to avoid regulatory arbitrage, where firms relocate to jurisdictions with weaker oversight. The work of [16] on leakage-safe benchmarks provides a template for regulatory evaluation of models across jurisdictions.

Finally, the fairness and equity of such systems must be continuously monitored. As models become more powerful, the risk of digital redlining, where certain firms are systematically disadvantaged by algorithmic predictions, increases. Policymakers should establish independent auditing bodies and require that any model used for capital allocation or supervisory ratings be subject to public accountability.

8. Conclusion

This paper has presented a comprehensive framework for multi-modal financial distress prediction that fuses price dynamics, macroeconomic signals, and corporate disclosures. We have argued that no single modality is sufficient to capture the complex, multi-faceted nature of financial distress, especially during periods of systemic stress. The proposed architecture addresses the temporal and structural heterogeneities of the three signal families through attention-based fusion and leakage-aware evaluation. We have highlighted the critical trade-offs between predictive accuracy, interpretability, robustness, and fairness, and we have outlined governance mechanisms to ensure responsible deployment. The integration of recent advances in benchmark design, stress indices, and attention models offers a path toward more credible and sustainable early warning systems. As financial markets become increasingly data-rich and interconnected, the multi-modal approach represents not just a technical improvement but a fundamental rethinking of how we model and manage financial risk.

References

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

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

3. Liu, T. (2026). PCA-APT Stress Index for Market Drawdowns.

4. Campbell, J. Y., Hilscher, J., & Szilagyi, J. (2008). In search of distress risk. Journal of Finance, 63(6), 2899–2939.

5. Bharath, S. T., & Shumway, T. (2008). Forecasting default with the Merton distance to default model. The Review of Financial Studies, 21(3), 1339–1369.

6. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

7. Sezer, O. B., Gudelek, M. U., & Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. Applied Soft Computing, 90, 106181.

8. Borio, C., & Lowe, P. (2002). Asset prices, financial and monetary stability: Exploring the nexus. BIS Working Papers, 114.

9. Liu, T. (2026). PCA-APT Stress Index for Market Drawdowns.

10. Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. The Journal of Finance, 66(1), 35–65.

11. Liu, T. (2022, December). Financial Constraint'Impact on Firms' ESG Rating Based on Chinese Stock Market. In 2022 4th International Conference on Economic Management and Cultural Industry (ICEMCI 2022) (pp. 1085-1095). Atlantis Press.

12. Zhang, Y., Xu, H., & Li, J. (2024). Multi-modal financial distress prediction: A survey. Expert Systems with Applications, 238, 122078.

13. Xue, P., & Ye, Y. (2026). Attention-enhanced reinforcement learning for dynamic portfolio optimization. Intelligent Systems with Applications, 200622.

14. Liu, T. (2026). PCA-APT Stress Index for Market Drawdowns.

15. Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. Journal of Machine Learning Research, 3, 1157–1182.

16. Liu, T. (2026). Leakage-Safe Benchmark Design for Market-Stress Early Warning: An Economically Credible Evaluation.

17. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15(1), 1929–1958.

18. Cui, Z., Chen, W., & Chen, Y. (2016). Multi-scale convolutional neural networks for time series classification. arXiv preprint arXiv:1603.06995.

19. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.

20. Liu, T. (2026). A Comparative Study of Transformer-Based and Classical Models for Financial Time-Series Forecasting. Journal of Risk and Financial Management, 19(3), 203.

21. Lample, G., & Conneau, A. (2019). Cross-lingual language model pretraining. Advances in Neural Information Processing Systems, 32.

22. Chen, D., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794.

23. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning deep features for discriminative localization. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2921–2929.

24. Li, J., & Xiong, Y. (2023). Causal discovery in financial time series: A review. Quantitative Finance, 23(5), 789–810.

25. Athey, S., & Imbens, G. W. (2016). Recursive partitioning for heterogeneous causal effects. Proceedings of the National Academy of Sciences, 113(27), 7353–7360.