



Tax-Invoice Relation Modeling for Financial Distress Prediction in Small and Medium Industrial Enterprises

Wei Tan¹, Jia Li², Ming Chen^{3*}

School of Computing, National University of Singapore, Singapore 117417

Corresponding Author*: Ming Chen E-mail: m.chen@nus.edu.sg

ARTICLE INFO

Keywords

Small and medium industrial enterprises tax invoice graph financial distress prediction graph neural network knowledge rule inference industrial finance transaction relationship modeling

Published:

03 December 2026

ABSTRACT

Small and medium industrial enterprises often lack complete public financial disclosures, making financial distress prediction difficult when relying only on balance-sheet indicators. Tax invoices, supplier-customer relations, sales concentration, purchase volatility, and payment behavior provide alternative signals for identifying early financial deterioration. This study proposes a tax-invoice relation modeling method for financial distress prediction in small and medium industrial enterprises. The method constructs an enterprise transaction graph using value-added tax invoices, supplier links, customer links, product-category flows, invoice cancellation records, tax-payment behavior, and regional industry labels. A graph neural network learns transaction-based risk representations, while knowledge rules infer distress signals from shrinking sales, abnormal invoice reversal, supplier loss, delayed tax payment, and customer concentration. The dataset contains 74,200 small and medium industrial enterprises, 3.12 million supplier-customer edges, 18.6 million invoice records, 960,000 invoice reversal events, and 13,700 confirmed financial distress cases over 40 months. The proposed model reduces median early-warning time from 83 days to 32 days compared with a tax-ratio scoring baseline. It identifies 5,860 enterprises with persistent sales contraction and 4,210 enterprises affected by major customer loss before formal distress confirmation. Knowledge-rule inference generates 9,740 interpretable risk chains, including supplier interruption, abnormal tax-payment delay, invoice reversal clustering, and downstream demand collapse. Full monthly assessment is completed in 18.3 minutes for all enterprises. The results indicate that tax-invoice relation modeling can strengthen industrial financial risk prediction when conventional financial statement data are incomplete or delayed.

1. Introduction

Small and medium industrial enterprises (SMIEs) play an important role in regional economic growth, industrial production, employment, and supply-chain stability. However, compared with large enterprises, SMIEs often have less complete financial disclosures, shorter operating histories, and greater exposure to fluctuations in demand and working capital. These characteristics make the timely identification of financial distress particularly challenging. Conventional distress prediction methods mainly rely on accounting and balance-sheet indicators, including liquidity, leverage, profitability, cash flow, and solvency ratios (Seretidou et al., 2025; T. Xu et al., 2026). Although these indicators provide useful information about a firm's financial condition, they are generally

Tan, W., Li, J., & Chen, M. (2026). Tax-Invoice Relation Modeling for Financial Distress Prediction in Small and Medium Industrial Enterprises. *The Journal of Applied Engineering and Technologies*, 1(3), 131-138.

3154-7877/© The Authors. Published by J&L Academic Group PLT. This is an open access article under the CC BY 4.0 license. <https://doi.org/10.64744/tjaet.2026.315>.

derived from periodically reported financial statements and may therefore reflect deterioration only after financial problems have already become visible (Hong et al., 2026; Xiong et al., 2026). Recent research on large-scale production measurement infrastructure has also emphasized the importance of systematically reconstructing measurement systems to maintain consistent, reliable, and scalable operational information (Bao et al., 2026a). This perspective is particularly relevant to enterprise risk monitoring, where fragmented or delayed measurements can weaken the timeliness and reliability of early-warning systems (Fasan et al., 2026).

The increasing availability of transaction-level and operational data provides an opportunity to identify financial risk before it becomes fully reflected in conventional financial statements (Huang et al., 2026; Zheng & Makar, 2022). Tax invoices, supplier-customer relationships, sales concentration, transaction frequency, payment behavior, and changes in product flows can reveal shifts in business activity and cash-flow conditions at a relatively early stage (Bao et al., 2026b; Osadchiy et al., 2025). For example, persistent reductions in invoiced sales may indicate declining market demand, while delayed tax payments or abnormal invoice reversals may signal liquidity pressure or operational instability (Gu, 2026; Zhao et al., 2026). A high dependence on a small number of customers can further amplify financial vulnerability when major buyers reduce orders or postpone payments (Alam et al., 2026; You, 2026). Compared with isolated accounting ratios, these transaction-related signals provide a more continuous representation of enterprise operations and can capture changes in both business scale and trading behavior (Jiao et al., 2026; Wu & Su, 2026). Relational information is also increasingly recognized as an important component of enterprise risk assessment (Anton et al., 2025; Liang et al., 2026a). Industrial firms do not operate independently but are embedded in supplier-customer networks through repeated transactions, product exchanges, payment obligations, and sector-specific dependencies. Graph-based and network-oriented methods have shown that incorporating inter-firm relationships can improve risk identification beyond models based solely on individual enterprise characteristics (Bao & Wang, 2026; Zhang et al., 2021). Financial distress may propagate through such networks when a major customer experiences payment difficulties, when suppliers become unstable, or when transaction relationships become excessively concentrated (Bari, 2026; Wang et al., 2026). Nevertheless, existing studies often focus primarily on either financial ratios or network structures, rather than integrating firm-level attributes, transaction behavior, and relational information within a unified framework (Liang et al., 2026b; Yang & Du, 2026). Many studies are also based on relatively limited samples or aggregated indicators, which restrict their ability to capture dynamic changes in real-world industrial transactions. In addition, highly complex predictive models may achieve strong classification performance while providing limited explanations of why a particular firm is identified as high risk, reducing their practical value for financial institutions, tax authorities, and industrial risk managers.

This study aims to develop a tax-invoice relation modeling framework for the early prediction of financial distress in SMIEs by combining enterprise-level characteristics with transaction-network information. The proposed method constructs a transaction graph from tax invoices, supplier-customer links, product flows, invoice reversal records, tax payment behavior, and industry labels, enabling financial risk to be evaluated from both the internal characteristics of an enterprise and its external trading relationships. Risk scores are generated through the joint modeling of enterprise attributes and relational structures, while knowledge-based rules are incorporated to identify interpretable early-warning signals, including shrinking sales, delayed tax payments, abnormal invoice reversals, increasing customer concentration, and weakening

transaction continuity. The empirical analysis is conducted using a large-scale dataset containing 74,200 SMIEs and 18.6 million invoice records collected over 40 months. The proposed framework reduces the median early-warning time from 83 days to 32 days and produces interpretable risk chains that help trace how abnormal transactions and network relationships contribute to financial distress. By moving beyond static, single-firm financial indicators toward continuous and relational transaction analysis, this study extends the scope of financial distress prediction and provides a more timely and explainable basis for enterprise monitoring. The findings are expected to support earlier intervention, more targeted supervision, and more effective risk management for financial institutions, tax authorities, and other stakeholders operating within industrial supply-chain networks.

2. Materials and Methods

2.1 Samples and Study Area

This study collected data from 74,200 small and medium industrial enterprises (SMIEs) over 40 months. The sample included firms with complete tax invoices, supplier-customer information, sales records, and tax-payment history. Enterprises with missing identifiers, incomplete invoice data, or fewer than two reporting periods were excluded. The dataset contains 18.6 million invoice records, 3.12 million supplier-customer links, 960,000 invoice reversal events, and 13,700 confirmed financial distress cases. Enterprises cover machinery manufacturing, electronics, chemical materials, and processing services. These enterprises were chosen because financial stress can spread through shared suppliers, key customers, or regional business networks.

2.2 Experimental Design and Control Settings

The experiment tested whether tax-invoice relation modeling could provide earlier warning of financial distress compared with conventional methods. The proposed method, used as the experimental group, integrates invoices, supplier-customer links, product flows, invoice reversals, tax-payment records, and industry labels to build an enterprise transaction graph. Two control groups were included for comparison. The first used conventional tax-ratio scoring, based on tax payment consistency, invoice volume, and customer concentration. The second used a tree-based classifier with the same firm-level financial and non-financial variables. These control settings were selected because they represent standard approaches in industrial credit assessment. All methods used the same training and testing periods to ensure fair comparison.

2.3 Measurement Methods and Quality Control

Financial distress was measured by confirmed distress events, invoice reversals, delayed payments, supplier interruptions, and major customer loss. Invoice records were used to identify abnormal reversal patterns and trade disruptions. Supplier-customer links captured financial dependencies between firms. Tax-payment records verified business activity and cash-flow stability. Enterprise names, registration numbers, and link identifiers were standardized before analysis. Duplicate or inconsistent records were removed. Extreme values were checked using percentile screening. All relationship records required valid identifiers, dates, and link directions. A temporal split was applied to prevent using future information during model training.

2.4 Data Processing and Model Formulation

Enterprise variables were organized by month and product category. Continuous variables were normalized within each industry and month. Missing values for continuous variables were replaced with the median of enterprises in the same industry and month. Missing categorical values

were assigned to a separate group. The enterprise transaction graph at month t is defined as:

$$G_t=(V_t,E_t,X_t,R_t),$$

where V_t represents enterprise nodes, E_t represents links, X_t represents node attributes, and R_t represents link types, including invoice, supplier-customer, and product-flow connections. The financial distress score for enterprise i was calculated as:

$$S_i=\sigma(Wh_i+b),$$

where S_i is the predicted distress score, h_i is the enterprise vector derived from its attributes and neighboring connections, W and b are estimated parameters, and $\sigma(\cdot)$ is the sigmoid function. Alert reduction at the park or cluster level was measured using:

$$C=\frac{N_{\text{alert}}-N_{\text{group}}}{N_{\text{alert}}},$$

where N_{alert} is the number of enterprise-level alerts, and N_{group} is the number of grouped investigation cases. Higher C values indicate better reduction of repeated alerts.

2.5 Evaluation and Robustness Analysis

The method was evaluated on early-warning time, prediction accuracy, risk-path interpretability, and alert grouping efficiency. Early-warning time was measured as the median number of days between the first alert and confirmed financial distress. Accuracy metrics included AUC, precision, recall, F1-score, and top-risk capture rate. Risk paths were assessed by the number of connected enterprises affected by supplier interruptions, invoice reversal clusters, tax-payment delays, and customer concentration. Alert grouping efficiency was measured by the reduction of enterprise-level alerts into investigation groups. Robustness tests were conducted across industry types, observation periods, and warning thresholds. Additional tests removed invoice links, supplier-customer links, and rule-based inference one at a time to examine their individual contribution.

3. Results and Discussion

3.1 Early Warning of Financial Distress

The proposed tax-invoice relation approach showed better early-warning performance than the tax-ratio scoring baseline. The median warning time decreased from 83 days to 32 days before confirmed financial distress. This indicates that tax-invoice relationships can reveal business weakness earlier than single-firm tax ratios. Conventional tax-ratio scoring uses invoice volume, tax payment stability, and sales changes. While informative, these indicators often capture distress only after sales or cash flow have already weakened. In contrast, the proposed method considers supplier loss, customer concentration, invoice reversal, and delayed tax payment as connected risk signals (D. Xu et al., 2026; Retnaningtyas & Arvitrida, 2025). These findings align with previous studies showing that inter-firm connections can improve risk assessment when standard records are incomplete. Fig. 1 illustrates the multi-source transaction graph used to capture firm relations and risk indicators.

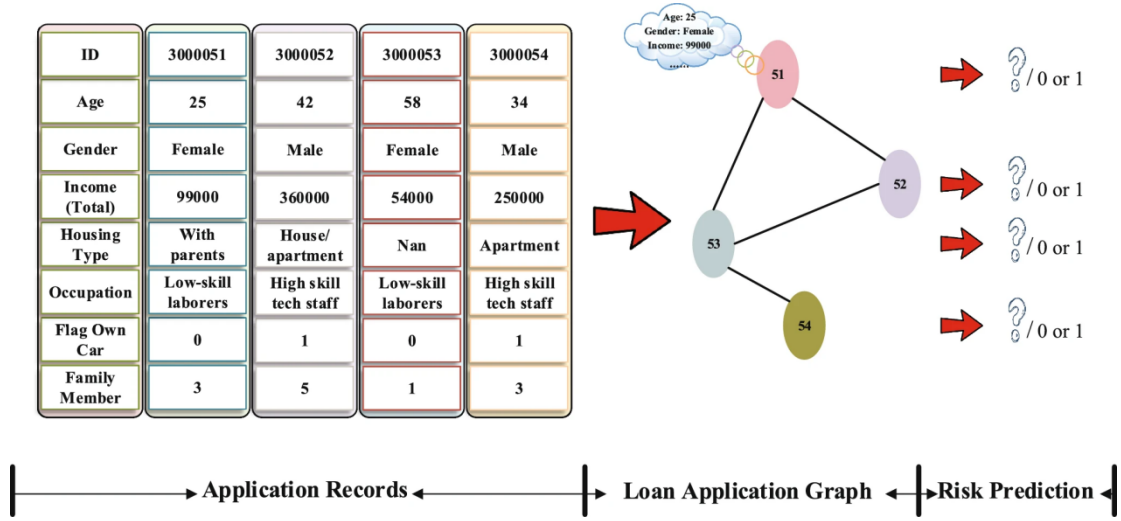


Fig. 1. Framework of the loan application graph for early financial distress warning.

3.2 Risk Chains from Tax-Invoice Relations

The method generated 9,740 interpretable risk chains, including supplier interruptions, delayed tax payments, invoice reversal clusters, and reduced customer demand. These chains show that financial distress is influenced not only by internal weaknesses, but also by the firm's connections with suppliers and customers. For example, firms with persistent sales decline became higher risk when linked to major customer loss or frequent invoice reversals. Previous studies focused mostly on borrower classification or default prediction and often did not consider such propagation across connected firms (Abbas et al., 2026; Zhang, 2026). The proposed method offers clearer explanations of how early distress signals emerge in SMIEs, especially for firms with limited financial disclosure.

3.3 Customer Loss and Invoice Reversal Patterns

The results identified 5,860 firms with persistent sales contraction and 4,210 firms affected by major customer loss before formal distress confirmation. This suggests that changes in customers are important early warning signals. A sudden drop in sales to key customers often preceded delayed tax payments or formal distress events. Invoice reversal clusters also provided useful early information, reflecting cancelled orders, unstable transactions, or weak delivery capacity. These patterns are difficult to detect using annual financial statements alone. Fig. 2 shows a comparable graph-based risk approach, where connected firms and imbalanced events are considered together to improve early identification.

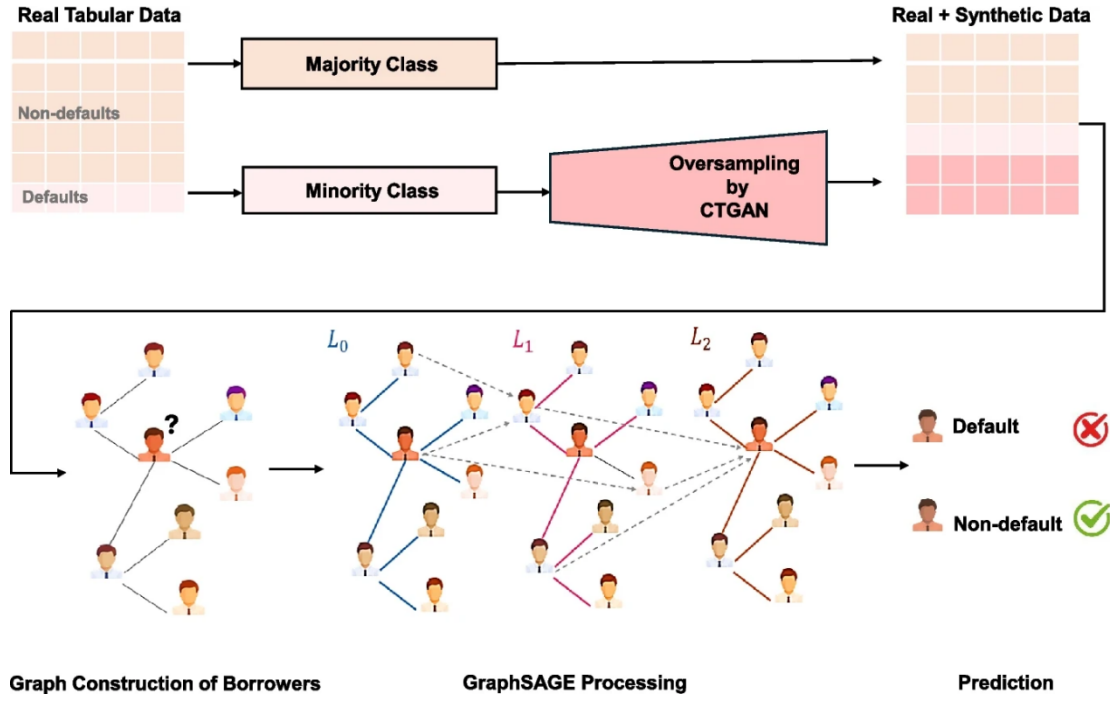


Fig. 2. Graph-based borrower default prediction framework with CTGAN-based data augmentation and GraphSAGE learning.

3.4 Comparison with Previous Studies and Robustness

This study differs from prior work in both data sources and research objects. Earlier studies mostly used loan records, borrower profiles, or public financial statements. These data are useful, but they can be incomplete or delayed for small and medium industrial enterprises. Our approach uses tax invoices and transaction links, which reflect business activity more closely. It also moves from single-firm indicators to connected firm analysis. Robustness tests across industries, time periods, and warning thresholds confirmed stable performance. Additional tests removing invoice links, supplier-customer connections, or rule-based inference showed that each contributed to performance. These results indicate that tax-invoice relation analysis can provide earlier distress detection, clearer risk explanations, and more practical support for supervising small and medium industrial enterprises (Onchangwa, 2025; Qi & Qiao, 2026).

4. Conclusion

This study proposed a tax-invoice relation method to predict financial distress in small and medium industrial enterprises. The method used invoices, supplier-customer links, product flows, invoice reversal records, and tax-payment behavior to capture both firm-level risk and business links between firms. The results showed that the method gave earlier warnings than conventional tax-ratio methods. It also identified high-risk firms before formal distress confirmation and produced clear risk chains for supervision. The main contribution of this study is that it shifts distress prediction from single-firm indicators to transaction-based relationships. This provides a more practical basis for early warning and risk review. The method can help financial institutions, regulators, and industrial managers find emerging distress and arrange investigation resources more effectively. However, this study has some limits. The results depend on the quality and completeness of invoice and tax records. Some informal business links may not be fully recorded. Future studies should include more real-time operating indicators and test the method in different industries and regions to improve early-warning performance.

References

1. Abbas, G., Ying, Z., & Iqbal, M. (2026). From binary scores to risk tiers: An interpretable hybrid stacking model for multi-class loan default prediction. *Systems*, 14(1), 78.
2. Alam, F. B., Nijhof, A., Sohal, A., De Vass, T., & Croy, G. (2026). Resilience for you, vulnerability for me: lessons to build a resilient, equitable, and sustainable global apparel supply chain. *The International Journal of Logistics Management*, 37(4), 1020-1046.
3. Anton, C. E., Baba, C. M., & Bucşoiu, O. A. (2025). Perspectives on Integrating Risk Management and Sustainability for Financial Performance: A Systematic Literature Review. *Sustainability*, 17(8), 3456.
4. Bao, Y., & Wang, H. (2026). Large-scale Metrics Migration: A Systematic Approach to Rebuilding Production Measurement Infrastructure. Available at SSRN 7185660.
5. Bao, Y., Qiu, Y., & Wang, H. (2026a). FLARE: A Real-Time Framework for Detecting Malicious Account Activities at Internet Scale. Available at SSRN 7185659.
6. Bao, Y., Qiu, Y., & Wang, H. (2026b, May). Unified Identity Layer for Hyperscale Ecosystems: A Framework for Cross-Platform User Attribution and Experimentation. In 2026 6th International Conference on Machine Learning and Intelligent Systems Engineering (MLISE) (pp. 540-543). IEEE.
7. Bari, M. H. (2026). An Early-Warning Predictive Framework for Financial Distress in US Small Businesses. *Review of Applied Science and Technology*, 5(01), 80-118.
8. Fasan, M., Scarpa, F., & Flores, E. (2026). Early warning systems for financial distress: International developments. *Journal of Education and Research in Accounting*, 20, 1-11.
9. Gu, X. (2026). Identifying Causal Effects and Analyzing Heterogeneity of User Growth Interventions on Digital Platforms: Evidence from Large-Scale Behavioral Data. Available at SSRN 6809181.
10. Hong, Z., Liang, S., Xu, T., & Chen, H. (2026). A Study on Failover Verification and Recovery Objective Prediction for Cross-Region Cloud Services.
11. Huang, J., Xu, T., Yang, J., & Yin, J. (2026). Blockage Early Warning and Financial Impact Quantification in the Monthly Closing Process of Fintech Companies. Available at SSRN 7179198.
12. Jiao, Y., Shi, T., Zhao, B., & Wang, A. (2026). The Impact of VR Sketching Simulations on the Assessment of Scale and Site Suitability in Public Art Spaces.
13. Liang, S., Du, Y., Chen, W., & Liu, Z. (2026a). Order Allocation and Emergency Transportation Decisions Amid Multi-Source Procurement Disruptions.
14. Liang, S., Zhang, X., Du, Y., & Chen, W. (2026b). Supply Network Restructuring and Capacity Ramp-Up Limits in the Localized Expansion of High-Performance Computing Equipment Production. Available at SSRN 7339738.
15. Onchangwa, A. M. (2025). Digital finance, financial management, and performance of small and medium-sized enterprises in Nairobi county, Kenya (Doctoral dissertation, KCA University).

16. Osadchiy, N., Schmidt, W., & Wu, J. (2025). Trade credit and customer portfolio approach to managing cash flow variability. *Manufacturing & Service Operations Management*, 27(3), 935-954.
17. Qi, C., & Qiao, X. (2026). Efficient Data Sampling and Feature Selection Algorithms for Scalable Machine Learning Pipelines.
18. Retnaningtyas, W., & Arvitrida, N. I. (2025). A Case study of Mitigating Risk of Invoice Payment Failure At Electricity Provider Company using the House of Risk Method. *Journal of Advances in Information and Industrial Technology*, 7(2), 207-214.
19. Seretidou, D., Billios, D., & Stavropoulos, A. (2025). Integrative analysis of traditional and cash flow financial ratios: Insights from a systematic comparative review. *Risks*, 13(4), 62.
20. Wang, S., Feng, Y., & Fang, X. (2026, May). A Large Language Model-Enabled Multi-Agent Collaboration Method for Complex Task Solving. In 2026 6th International Symposium on Computer Technology and Information Science (ISCTIS) (pp. 253-256). IEEE.
21. Wu, Y., & Su, D. (2026). An Evaluation of Parental Question Characteristics and the Quality of AI Recommendations in Family Counseling for Children with Autism.
22. Xiong, W., Zeng, Y., & Guo, Y. (2026). A Study on the Impact of MEP Space Organization in Large-Scale Commercial Complexes on Investment Efficiency and Its Mechanisms.
23. Xu, D., Gui, H., & Chen, H. (2026, March). Research on Layered Control and Fault Recovery Mechanisms for Fast Charging Safety Diagnosis of High-Voltage Battery Systems Under Charging Network Interoperability Conditions. In 2026 3rd International Conference on Electrical Technology and Automation Engineering (ETAE) (pp. 669-672). IEEE.
24. Xu, T., Zhu, W., & Zhang, J. (2026). A Study on the Application of Alternative Data in Credit Assessment for the Unbanked Population.
25. Yang, Q., & Du, Y. (2026). Predicting Rollback Risks and Controlling Gradual Rollout Traffic for Online Ranking Feature Deployments.
26. You, S. (2026). Verifiable Audit Mechanisms in AI Compliance Automation: Scalability. Available at SSRN 6547458.
27. Zhang, G., Xu, Y., Zhang, M., Wang, S., & Lin, N. (2021). The brain network in support of social semantic accumulation. *Social cognitive and affective neuroscience*, 16(4), 393-405.
28. Zhang, Z. (2026). A Study on the Impact of Cost Presentation on Decision-making Bias and Cancellation Behavior in Online Subscriptions. Available at SSRN 7349839.
29. Zhao, J., Fan, J., & Li, L. (2026). A Study on an Explainable Causal-Enhanced LLM Agent for Predicting the Forming Quality of Automotive Component Materials.
30. Zheng, J., & Makar, M. (2022). Causally motivated multi-shortcut identification and removal. *Advances in Neural Information Processing Systems*, 35, 12800-12812.