Big Data and Machine Learning Methods in Financial Risk Prediction

Yangguang Xu*

Xi'an Jiaotong-Liverpool University, Suzhou, China

*17374848791@163.com

Abstract

Financial risk prediction has become increasingly critical as modern financial systems face rising uncertainty, complex market dynamics, and rapidly expanding data sources. Despite progress in traditional risk assessment, recent crises have revealed persistent limitations in capturing nonlinear and emergent risk patterns. This study aims to systematically examine how big data and machine learning methods are reshaping financial risk prediction and to identify the key trade-offs between predictive performance and model interpretability. We review major methodological developments across supervised and unsupervised learning, deep neural networks, graph-based models, multimodal architectures, and emerging large language model–driven frameworks. The findings show that advanced models, especially deep, graph-based, and multimodal approaches, consistently outperform traditional statistical techniques, yet often do so at the expense of transparency, giving rise to a growing focus on explainable and human-centered financial AI. We also highlight emerging trends such as multimodal financial foundation models, privacy-preserving federated learning, and inherently interpretable architectures. This review provides guidance for researchers and practitioners seeking to build financial risk models that are both accurate and trustworthy in real-world applications.

Keywords

Financial risk prediction, big data, machine learning, explainable AI / XAI, multimodal models.

1. Introduction

Financial risk prediction constitutes an essential component of modern financial systems, as its accuracy directly influences the stability and profitability of financial institutions. Whether the focus lies on credit risk, market risk, or operational risk, inaccurate forecasts can trigger severe economic consequences. For instance, the global financial crisis of 2007–2008 exposed the limitations of traditional risk assessment approaches in handling complex derivatives and cross-market exposures [1, 2], while the 2023 collapse of Silicon Valley Bank again highlighted that shortcomings in modeling and risk management can induce systemic disruptions [3]. Thus, developing more precise and reliable financial risk prediction models has become a critical task for ensuring the resilience of the financial system. Within this context, the full-scale digitalization of financial services and the rise of internet finance have significantly transformed the data foundation and modeling environment of risk prediction. Financial institutions now have access to massive, multi-source data, including transaction histories, financial statements, news articles, and social media content. Such big-data environments offer unprecedented breadth and depth for risk identification, yet they also increase modeling complexity and computational requirements [4]. In response, machine learning and deep learning techniques have become increasingly important tools, owing to their strong feature-

extraction capabilities and proficiency in learning nonlinear patterns. These methods not only complement traditional statistical models but also demonstrate superior predictive performance across numerous tasks. Despite their advantages in predictive accuracy, machine learning and deep learning models face persistent challenges related to interpretability, class imbalance, and privacy concerns. In financial applications, characterized by stringent regulatory oversight and sensitivity, models must provide not only high accuracy but also transparency acceptable to regulators, clients, and investors. Consequently, reviewing existing methodologies, clarifying the applicability of different modeling approaches, and examining the trade-off between predictive performance and interpretability are essential for advancing research and practice in this field. Based on these challenges, this study addresses three central research questions: 1. What is the current landscape of big data and machine learning applications in financial risk prediction? 2. How can predictive performance and model interpretability be effectively balanced? 3. What are the potential directions for future research and practice?

2. Machine Learning Applications in Financial Risk Prediction

To accommodate the diversity and complexity of financial risk data, contemporary risk-prediction frameworks employ a broad range of machine learning techniques. These approaches span traditional statistical models to advanced deep learning architectures, forming a multi-layered and multi-modal modeling ecosystem.

2.1. Traditional Supervised Learning: Foundational Models and Performance Benchmarks

Supervised learning plays a foundational role in financial risk prediction. Classical algorithms such as logistic regression, support vector machines, and decision trees have been extensively used in default prediction due to their transparency and computational efficiency. Ensemble methods such as random forests and XGBoost, which offer strong nonlinear approximation capabilities, are now predominant for modeling structured data, including financial ratios and market indicators [5]. However, the performance of such models depends heavily on high-quality feature engineering, and their interpretability regarding complex risk drivers remains limited [5].

2.2. Unsupervised and Semi-Supervised Learning: Uncovering Latent Structures

Unsupervised and semi-supervised learning complement supervised methods by enabling analysis of unlabeled or partially labeled data. Clustering techniques, for example, can identify groups of firms with similar risk profiles, while autoencoders learn robust latent representations through data reconstruction, supporting anomaly detection tasks [6]. Nonetheless, these methods require careful calibration when applied to default prediction and often lack direct business interpretability, necessitating the integration of domain expertise.

2.3. Deep Learning: Automatic Extraction of Complex Patterns

To overcome the limitations of manual feature engineering and capture deeper nonlinear relationships, deep learning models have been increasingly adopted. Convolutional neural networks (CNNs) extract local interaction patterns from matrix-structured financial indicators, whereas recurrent neural networks (LSTM/GRU) are effective for modeling long-term dependencies in financial time series. Hybrid architectures combining CNNs and RNNs, along with Transformer models based on attention mechanisms, have further demonstrated strong potential in risk prediction tasks [5]. Despite these performance gains, concerns persist

regarding the complexity and black-box nature of deep models, which complicate interpretability and reliability assessments [5, 7].

2.4. Graph Neural Networks: Modeling Risk Propagation in Relational Structures

Graph neural networks (GNNs) are well-suited for analyzing relational financial data, such as ownership structures and transaction networks. Empirical studies show that GNNs effectively capture such graph-structured relationships and achieve strong predictive accuracy in risk-related tasks [8]. Applications include modeling borrower–guarantor networks for bankruptcy prediction and encoding intra-firm accounting relationships to analyze risk contagion [5, 8]. By shifting the analytical perspective from individual entities to interconnected systems, GNNs offer powerful tools for systemic-risk modeling.

2.5. Multimodal and Hybrid Methods: Integrating Heterogeneous Information

Multimodal and hybrid approaches represent the frontier of current research, integrating numerical, textual, audio, and other data types. A typical strategy involves combining structured financial indicators with unstructured textual data from annual reports or news, using BERT-based language encoders in conjunction with numerical models to enhance credit-rating prediction accuracy [5, 9]. More conceptually, multimodal financial foundation models (MFFMs) aim to construct unified platforms capable of processing tabular, textual, and audiovisual inputs [9]. While multimodal fusion broadens available informational signals, it also increases the difficulty of data alignment, annotation, and model training.

3. The Trade-off between Performance and Interpretability

In financial applications, interpretability is as vital as predictive accuracy. A widely recognized trade-off exists: simple white-box models (e.g., linear regression, shallow decision trees) are more interpretable but struggle to capture complex nonlinear relationships, whereas black-box models (e.g., deep neural networks, large language models) often yield higher predictive accuracy at the cost of reduced transparency [6]. For instance, multichannel deep models can substantially improve firm-level risk prediction accuracy, yet their complex architecture and large parameter space make direct interpretation challenging [5]. Similarly, hybrid CNN–LSTM or Transformer-based architectures usually offer superior predictive performance with limited transparency [5]. Such opacity is particularly problematic in finance, where regulators and risk managers require auditability and traceability, making blind reliance on non-interpretable models unacceptable. Several reviews and empirical studies highlight these issues. Yeo et al. observe that neural networks and deep learning are often regarded as black-box methods in finance, while also noting that white-box models may fail to achieve the required performance in certain settings [6]. Conversely, findings by Rudin and Radin indicate that well-designed interpretable models can match complex models in several financial tasks [7], suggesting that transparent models should be prioritized when feasible. To bridge the performance-interpretability gap, multiple categories of explainable AI (XAI) techniques have been proposed, including model-agnostic tools (e.g., SHAP, LIME, counterfactual explanations), inherently interpretable architectures, and surrogate models that approximate black-box behavior. Some studies apply attention heatmaps and Shapley-value analyses to quantify the importance of different data modalities and time steps [10]. Recent reviews also explore integrating large language model (LLM) outputs with interpretable structures such as decision trees or token-level explanations [11]. Prototype-based explanations and attention masks further provide human-readable evidence on representative financial features [5]. Despite these advances, achieving an optimal balance between accuracy and interpretability remains a central

challenge: enhancing interpretability often requires compromises in model complexity or performance [6].

4. Future Directions

Recent technological trends are reshaping the paradigm of financial risk modeling. Compared with traditional approaches that rely primarily on structured financial ratios, future risk-prediction systems are expected to increasingly leverage multimodal fusion, large language models, explainable AI, and privacy-preserving collaborative learning frameworks.

4.1. Multimodal Data Fusion

As accessible financial signals continue to diversify, risk prediction is transitioning from single-source to multimodal modeling. Future frameworks will integrate structured indicators, financial text, news events, audio recordings (such as earnings calls), and even visual data within unified architectures to capture a more complete representation of firm-level risks. Multimodal financial foundation models (MFFMs) exemplify this trend by providing unified encoders for market data, financial documents, macroeconomic indicators, and alternative signals [9]. Empirical evidence shows that blending sentence-level embeddings from annual reports with financial ratios significantly improves financial-distress prediction [5]. However, multimodal fusion also introduces challenges related to data alignment, scarcity of labeled samples, and increased computational complexity.

4.2. Large Language Models in Finance

Domain-specific LLMs (e.g., BloombergGPT, FinGPT) and rapidly evolving general-purpose LLMs are reshaping financial text analysis. LLMs can extract fine-grained, context-rich risk signals from financial reports, news, and earnings-call transcripts while also generating human-readable explanations. Current approaches involve fine-tuning LLMs on financial corpora, adopting Transformer architectures capable of handling long documents, and integrating LLMs with traditional decision models [11]. Empirical findings show that LLM-based sentiment extraction often outperforms dictionary-based methods and yields improvements in credit-risk prediction [6, 11]. For example, GPT-3.5 offers more robust sentiment extraction than classical text-processing techniques when analyzing lengthy financial reports [11].

Nonetheless, LLMs face challenges such as hallucinations, difficulties in auditing generated content, high inference costs, and compliance risks under financial regulation [6, 9]. Future research may focus on controllable generation strategies, robustness enhancements, and hybrid architectures equipped with self-verification mechanisms.

4.3. Explainable Finance and Human-Centered AI

Given stringent regulatory requirements, interpretability is nearly as important as predictive performance in financial settings, driving the emergence of explainable finance (FinXAI). Current methods include inherently interpretable models such as rule sets, model-agnostic explanation techniques such as SHAP and LIME, and natural-language explanations generated by LLMs [6]. These approaches are increasingly combined with visualization interfaces to support case-level analysis. Persistent trade-offs remain: white-box models offer transparency but struggle in high-dimensional nonlinear scenarios; black-box models excel in performance but rely on approximation-based explanations that may not fully reflect underlying logic. The emerging trend shifts from post-hoc explanations toward architectural transparency, embedding interpretability directly into model design and incorporating domain knowledge from accounting and finance.

4.4. Federated and Privacy-Preserving Learning

Due to regulatory constraints, cross-institutional sharing of raw financial data is becoming increasingly difficult, making federated learning (FL) an appealing alternative. By performing local training with privacy-preserving model aggregation, FL enables institutions to collaboratively build risk-prediction models without sharing sensitive data [12]. Empirical work demonstrates its value in fraud detection, credit scoring, and supply-chain finance, helping alleviate the persistent issue of limited default data. Remaining challenges include heterogeneous data distributions, communication overhead, and vulnerability to adversarial attacks. Future developments may focus on encrypted aggregation, robustness enhancements, and standardized cross-bank protocols.

4.5. Additional Emerging Themes

Self-supervised learning (SSL) offers a promising method for reducing labeling costs by enabling models to extract general representations from large volumes of unlabeled time-series or textual data. Incorporating domain knowledge, such as regulatory logic or macroeconomic scenarios, via knowledge graphs represents another promising direction. Furthermore, fairness, robustness, and adversarial resilience (e.g., resistance to manipulated news events) are becoming increasingly critical in practical financial-risk systems.

5. Conclusion

Financial risk prediction is evolving from traditional models centered on financial ratios toward data-driven paradigms that integrate heterogeneous large-scale data sources. Modern approaches now span supervised and unsupervised learning, deep neural networks, and relational modeling techniques, yielding substantial empirical performance gains. However, improvements in accuracy often come at the cost of reduced transparency, elevating explainability to a central concern in both research and industry applications. A fundamental challenge persists: complex models excel at capturing nonlinear risk patterns yet offer limited interpretability, whereas interpretable models struggle to maintain comparable predictive accuracy in high-dimensional settings. Looking ahead, integrating multimodal data sources (text, audio, images) and combining them with large language model architectures is expected to enhance the expressiveness of risk-prediction models. In parallel, advances in interpretability methods and privacy-preserving learning, such as federated learning, will be essential for practical deployment. Overall, future research is likely to pursue hybrid solutions that balance predictive accuracy with transparency, enabling highly performant yet trustworthy and regulator-aligned financial AI systems.

References

[1] Jorion, P. (2009). Risk management lessons from the credit crisis. European Financial Management, 15(5), 923–933. https://doi.org/10.1111/j.1468-036X.2009.00507.x

[2] Degiannakis, S., Floros, C., & Livada, A. (2012). Evaluating value-at-risk models before and after the financial crisis of 2008: International evidence. Managerial Finance, 38(4), 436–452. https://doi.org/10.1108/03074351211207563

[3] Metrick, A. (2024). The failure of Silicon Valley Bank and the panic of 2023. Journal of Economic Perspectives, 38(1), 133–158. https://doi.org/10.1257/jep.38.1.133

[4] Dastile, X., Çelik, T., & Potsane, M. M. (2020). Statistical and machine learning models in credit scoring: A systematic literature survey. Applied Soft Computing, 91, 106263. https://doi.org/10.1016/j.asoc.2020.106263

[5] Zheng, H., Ma, Y., & Wang, J. (2025). Financial risk forecasting with RGCT-PreRisk: A relational graph and cross-temporal contrastive pretraining framework. Journal of King Saud University – Computer and Information Sciences. https://link.springer.com/article/10.1007/s44443-025-00166-4

[6] Yeo, W. J., Van Der Heever, W., Mao, R., et al. (2025). A comprehensive review on financial explainable AI. Artificial Intelligence Review, 58, 135–162. https://link.springer.com/article/10.1007/s10462-024-11077-7

[7] Rudin, C., & Radin, J. (2019). Why are we using black box models in AI when we don't need to? A lesson from finance. Harvard Data Science Review, 1(2). https://doi.org/10.1162/99608f92.5a8a3a3d

[8] Chen, Y., Sun, X., & Liu, P. (2025). A review on graph neural network methods in financial applications. Journal of Data Science, 22(2), 145–160. https://jds-online.org/journal/JDS/article/1279/info

[9] Zhang, L., Wu, F., & Yang, J. (2025). Multimodal financial foundation models (MFFMs): Progress, prospects, and challenges. arXiv preprint arXiv:2506.01973. https://arxiv.org/html/2506.01973v2

[10] Korangi, A., Hussain, B., & Alqahtani, H. (2021). A transformer-based model for default prediction in mid-cap corporate markets. arXiv preprint arXiv:2111.09902. https://arxiv.org/abs/2111.09902

[11] Li, J., Zhang, S., & He, L. (2025). Interpretable LLMs for credit risk: A systematic review and taxonomy. arXiv preprint arXiv:2506.04290. https://arxiv.org/html/2506.04290v2

[12] Al-Sharif, M., Khan, A., & Hussain, T. (2024). Federated machine learning in finance: A systematic review on technical architecture and financial applications. In Proceedings of Applied and Computational Engineering (pp. 16640–16648). https://www.ewadirect.com/proceedings/ace/article/view/16640