Proceedings of ICFTBA 2026 Symposium: Digital Government, Finance, Accountability, and Sustainable Economic Development
DOI: 10.54254/2754-1169/2026.AB38991

Applications of Artificial Intelligence and Machine Learning in Financial Risk Management: A Review of Recent Research

Wenze Guo

Individual researcher, Xiamen, China
guowenze0717@163.com

Abstract. Artificial intelligence (AI) and machine learning (ML) are now widely used in credit, market, operational, liquidity, and systemic risk management, largely because they can handle large-scale, multi-source, and nonlinear data. This literature review draws on recent studies and public cases and groups the evidence by these five risk categories. It compares reported methods, data sources, evaluation practices, and regulatory issues. Credit scoring and fraud detection have advanced fastest. Market risk models increasingly rely on deep learning for volatility and return prediction, while liquidity and systemic risk remain difficult to model because extreme events are rare. Random forests, gradient boosting, and neural networks often outperform linear models on public datasets, yet their advantage is less stable in live business settings and stressed markets. Data silos, class imbalance, model degradation, weak interpretability, privacy risks, adversarial attacks, and model homogeneity continue to limit deployment. More work is needed on explainable AI, privacy-preserving learning, stress testing, and cross-institutional data governance, and validation on proprietary data remains necessary before wider use.

Keywords: Financial risk management, credit risk, fraud detection, model explainability

1. Introduction

In recent years, with the rapid development of financial technology and the increasing application of machine learning in financial risk management, the operational model and risk landscape of the global financial industry are undergoing profound changes. Against this backdrop, it is difficult for traditional financial models to deal with the increasingly complex dynamics of the financial market and the risk environment. Many studies have widely discussed the possibility of applying artificial intelligence and machine learning in areas such as credit risk and market risk. Some local financial authorities in China have already implemented artificial intelligence. For instance, the "Financial Brain" platform in Wenzhou, Zhejiang Province, was jointly developed by the Wenzhou municipal government and Ant Financial. It created a regulatory technology system for "financial risk warning and financial services", achieving comprehensive monitoring of over 250,000 enterprises in Wenzhou and enhancing regulatory efficiency and expanding the scope of supervision. However, artificial intelligence still faces issues such as insufficient interpretability and poor data quality. This paper aims to systematically review the current application status of artificial intelligence and machine learning in financial risk management, including credit risk, market risk, operational risk,

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2754-1169/2026.AB38991

liquidity risk, and systemic risk. It also explores the main applications in these fields and discusses potential future research directions. The goal is to help relevant practitioners quickly understand the current situation in this field and to provide a reference for better integration of artificial intelligence into financial risk management.

2. The definitions and types of financial risks

2.1. Financial risk

Financial risks are generally regarded as potential losses that may occur when financial institutions or market participants engage in market activities due to market fluctuations, counterparty default, or internal operational errors. Currently, the financial risks frequently discussed in the academic community can be classified as credit risk, market risk, operational risk, liquidity risk, and systemic risk, which affects institutions in an interconnected manner [1]. Credit risk refers to the possibility of borrowers or counterparties failing to fulfill their debt obligations as agreed, or experiencing a deterioration in their credit status, resulting in losses. Market risk generally refers to adverse changes in market prices, such as unfavorable changes in interest rates, exchange rates, stock prices, and commodity prices. Operational risk generally refers to risks caused by internal procedures, personnel, deficiencies in internal systems, and force majeure factors such as natural disasters. Liquidity risk generally refers to the risk that assets cannot be quickly converted into cash without a significant reduction in price, or the risk of being unable to raise funds at a reasonable cost to repay debts. Systemic risk generally refers to the risk that a single financial institution or market event, through contagion effects, leads to widespread functional impairment or cascading failures in the financial system.

2.2. Reasons for applying artificial intelligence and machine learning to financial risk

Most traditional management methods depend on analyzing financial statements and using credit ratings to control risks. However, these methods neglect the complexity of financial markets and have limited ability to identify complex nonlinear relationships. The growth in the scale and complexity of financial data has raised new demands for analytical tools. Modern financial data include not only structured transaction records and financial statements but also unstructured information such as text (news, announcements, and social media), images (bills and certificates), and high-frequency time series. Traditional reports and linear models are no longer sufficient, so machine learning provides new solutions [1]. Currently, machine learning has been applied in multiple financial scenarios such as credit scoring, fraud detection, investment, and risk prediction [2].

3. Applications of artificial intelligence and machine learning in different types of financial risk

3.1. Applications in credit risk

Credit risk refers to losses caused by borrowers or counterparties failing to repay on time. Traditional credit risk management usually relies on expert scoring and logistic regression. These methods are convenient for explaining decisions to regulators, but they have difficulty conducting multi-dimensional assessments of borrowers or counterparties, handling nonlinear relationships, and addressing the complexity of credit risk. Existing research mostly treats default prediction as a

DOI: 10.54254/2754-1169/2026.AB38991

classification problem. Montevechi et al. conducted a comprehensive review of credit risk modelling with machine learning and found that ensemble methods such as random forests have stronger discrimination ability than a single logistic regression on most datasets [3]. Later research further applied gradient boosting trees to personal and micro-loan default prediction and attempted to use neural networks to process cash flow, text, and other information, thereby supplementing multi-dimensional evaluation of customers [4]. Alternative data can enhance predictive power, but it also raises privacy issues and sample bias problems [5]. Wenzhou City has implemented big data-assisted credit evaluation and risk warning for supervision and decision-making support [6]. Some studies have pointed out that machine learning has been used for credit rating and default discrimination, with the focus still being on improving identification ability rather than replacing the loan review system [7].

3.2. Applications in market risk

Market risk mainly arises from adverse changes in asset prices, interest rates, and exchange rates. Traditional methods for managing market risk mainly rely on volatility models and value at risk to describe complex situations. The review literature mentions that common methods of applying machine learning to market risk and investment portfolios include predicting returns and volatility, as well as using models to assist in constructing risk measurements [1, 2]. Existing studies have pointed out that deep learning methods such as LSTM and CNN are used to extract temporal features from price sequences for risk prediction, these applications depend more on continuous market data and are more sensitive to out-of-sample stability and extreme market conditions [5]. However, improving prediction accuracy does not equal providing reliable regulatory capital or trading limits. The setting of regulatory capital and trading limits also depends on distribution assumptions, backtesting, and stress testing, and merely improving in-sample prediction accuracy is not sufficient [8].

3.3. Applications in operational risk and fraud detection

Operational risk arises from internal processes, personnel, system failures, or external events. Network attacks and fraud detection are the scenarios where machine learning is most widely applied in operational risk. Research often lists anti-fraud and abnormal transaction identification as high-frequency applications, and commonly uses supervised classification and unsupervised anomaly detection [1, 2]. Literature related to internet finance points out that account theft, fraudulent applications, and transaction fraud are scenarios where algorithms were implemented earlier, and deep models are used to identify anomalies from transaction records and device information, but interpretability and adversarial attacks and evasion remain issues [5, 9]. Chinese discussions also consider artificial intelligence as both a risk control tool and a new risk source: models can be used to identify suspicious transactions, and their own data leakage and algorithm vulnerabilities can also be transformed into operational and security risks [10, 11]. Therefore, when evaluating such systems, not only accuracy should be considered, but also whether false alarms exhaust manual review and whether fraud patterns shift after the rules are exposed.

3.4. Applications in liquidity risk and systemic risk

Liquidity risk refers to the risk of being unable to liquidate assets at a reasonable price in a timely manner or to raise funds at a reasonable cost to repay debts. Systemic risk, on the other hand, refers

DOI: 10.54254/2754-1169/2026.AB38991

to the situation where problems in a single institution or a single market spread throughout the entire financial system. Due to the rarity of extreme cases of these two types of risks, it is more difficult to directly label samples for these risks compared to individual loans or individual stocks. In the review literature, they are mostly discussed in the context of macro monitoring and portfolio management. The main role of machine learning is to identify the sources of stress in various market indicators [1, 2]. Existing studies also focus on integrating dispersed data into early warning systems [12]. The early warning system of the International Monetary Fund has adopted machine learning methods to identify early signals of systemic banking crises, providing a reference for macroprudential policies [1]. However, warning signals do not equal accurate predictions of systemic collapses. The true reliability of the model still needs to be tested in stress events.

4. Current challenges and limitations

4.1. Data quality and availability

The credit default and fraud records of financial institutions and platforms are subject to strict regulation, making it difficult for these entities to share data with one another. Most models can only be trained using their own internal data [1]. Current research heavily relies on a few publicly available datasets, such as credit approval data from Germany and Australia, Lending Club platform data, and European credit card transaction data from September 2013 [1, 4]. However, when the same algorithm is transferred to different datasets, model ranking performance often fluctuates. There are also significant differences between the actual customer groups and application characteristics of public datasets and those of individual institutions [4]. In the context of internet finance, the prediction accuracy of neural networks depends heavily on model architecture, sample features, and parameter configuration, and no universal algorithm can be applied to all platforms [9]. Moreover, the proportion of default and fraud samples in the overall population is extremely low, so non-default samples far outnumber default samples. Simple classifiers tend to ignore minority-class samples and misclassify them as the majority class [4, 9]. This data imbalance is an inherent characteristic of financial risk data and cannot be fully eliminated by simple resampling techniques [5]. At the same time, historical data may contain past decision biases, which can lead the model to mistakenly learn that those who did not receive loans in the past are at high risk of future default, thereby reinforcing existing screening biases.

4.2. Model risk and overfitting

Machine learning models essentially fit the distribution of historical data. Financial time series are highly noisy, and correlations and volatility change over time [1]. For instance, events, consumer behavior, default rates, and market expectations all change simultaneously. If a model trained during a stable period is still used, an adaptation lag will occur. Some institutions have acquired and resold housing through algorithms on a large scale but were misled by extreme events in judging asset growth and operational costs, ultimately resulting in project shutdown and large-scale asset impairment. This is the consequence of distribution drift while the model still operates according to old rules [10]. Moreover, the prediction results of neural networks are relatively sensitive to weight initialization and hyperparameter selection. Even changing the initialization parameters alone can lead to different conclusions [1]. In bank risk management, such problems are usually collectively referred to as "model risk", specifically manifested as poor interpretability, excessive sensitivity to input data, and overfitting on the training set [8]. Although deep learning and ensemble methods

DOI: 10.54254/2754-1169/2026.AB38991

typically outperform single models or traditional statistical methods on public datasets, this only represents their ranking ability on specific datasets and cannot be directly inferred to hold in different years or at different institutions [4]. Currently, there are still relatively few specialized studies on extreme risk scenarios in the literature related to internet finance [9].

4.3. Limited interpretability and regulation

When a loan application is rejected or an account is frozen, financial institutions usually need to explain the specific reasons for the decision to customers [1]. Traditional decision tree models classify based on a series of conditional judgments and can later roughly trace whether the decision was due to income level, number of past delinquencies, or other specific indicators. In contrast, neural networks with more layers usually only output a comprehensive score, making it difficult to clearly explain the specific underlying reasons [1, 8]. Multiple reviews in the field of credit risk have listed the "lack of transparency" of the model as one of the core difficulties [4]. The opacity of the decision rationale can lead to serious incidents in actual business. For example, in 2015, the Australian federal government introduced automated algorithms to recover unpaid taxes, but due to the lack of transparency in the algorithm's operation, a large number of incorrect bills were not corrected or compensated until many years later [10]. In the U.S. housing credit market, if a model unintentionally learns and internalizes historical regional discrimination patterns, rejected applicants may not even know why they were excluded and thus have difficulty making effective appeals [10].

4.4. Extreme events and synchronized failures

Extreme events such as financial crises are relatively rare in historical data, which leads to models being trained mostly on data from stable years. However, when a large number of institutions adopt highly similar data sources and model architectures, systemic risk of "coordinated misjudgment" can arise [10]. For example, during the Asian morning trading session on October 7, 2016, liquidity in the British pound was relatively thin, and automated sell orders and stop-loss orders overlapped, causing the exchange rate to drop sharply and then rebound. Ding and Wang used this example to illustrate the risk of algorithms generating synchronous responses to similar signals, and pointed out that this event was not directly driven by artificial intelligence [10]. Credit card anti-fraud systems are also vulnerable to specially designed adversarial examples [10]. Related reviews on internet finance also point out that financial risk prevention must pay close attention to extreme situations, but the coverage of existing research in this field is still insufficient [9]. The credit warning platform integrates scattered data into a visual dashboard, but this does not mean that it can accurately predict the next large-scale risk event [6].

5. Conclusion

This paper categorizes applications of artificial intelligence and machine learning in financial risk management by risk category. In credit risk, most existing studies treat default prediction as a classification problem, and ensemble methods and neural networks have improved discrimination on public datasets. However, ranking performance changes when the dataset is replaced, and alternative data may also raise privacy concerns and sample bias. In market risk, research mainly focuses on predicting returns and volatility and assisting risk measurement. These applications are more sensitive to out-of-sample performance and extreme market conditions. Higher prediction accuracy does not mean that regulatory capital or trading limits have become reliable. Applications in

DOI: 10.54254/2754-1169/2026.AB38991

operational risk and fraud detection emerged earlier, but the difficulties lie in the scarcity of fraud samples, false alarms consuming manual review capacity, and fraud tactics shifting after rules are exposed. Labeling samples for liquidity and systemic risk is more difficult. Machine learning can often identify signs of stress from multiple indicators, but warning signals cannot be equated with accurate predictions of the next large-scale collapse.

Based on the existing literature, the main limitations remain the limited sharing of default and fraud data among institutions and the significant differences between public datasets and actual customer populations. Models are fitted to the historical data distribution and may become ineffective when the market environment changes. Deep neural networks often output only a comprehensive score, making it difficult to explain the decision rationale when loans are rejected or accounts are frozen. Systems trained in stable periods may not cope with extreme market conditions or withstand the simultaneous use of similar models by multiple institutions. This review is based mainly on existing literature and public cases and lacks direct validation using real business data. Further research is needed to examine the stability of sample replacement in real environments, model failure under stress scenarios, and how to explain decision rationales to regulators and customers.

References

[1] Mashrur, A., Luo, W., Zaidi, N. A., & Robles-Kelly, A. (2020). Machine learning for financial risk management: A survey. *IEEE Access, 8*, 203203–203223. https://doi.org/10.1109/ACCESS.2020.3036322

[2] Nazareth, N., & Reddy, Y. V. R. (2023). Financial applications of machine learning: A literature review. *Expert Systems with Applications, 219*, 119640. https://doi.org/10.1016/j.eswa.2023.119640

[3] Montevechi, A. A., Miranda, R. de C., Medeiros, A. L., & Montevechi, J. A. B. (2024). Advancing credit risk modelling with machine learning: A comprehensive review of the state-of-the-art. *Engineering Applications of Artificial Intelligence, 137*, 109082. https://doi.org/10.1016/j.engappai.2024.109082

[4] Shi, S., Tse, R., Luo, W., D'Addona, S., & Pau, G. (2022). Machine learning-driven credit risk: A systemic review. *Neural Computing and Applications, 34*, 14327–14339. https://doi.org/10.1007/s00521-022-07472-2

[5] Peng, K., & Yan, G. (2021). A survey on deep learning for financial risk prediction. *Quantitative Finance and Economics, 5*(4), 716–737. https://doi.org/10.3934/QFE.2021032

[6] Jin, H. (2024). *Research on Wenzhou municipal government financial supervision in the era of artificial intelligence: A case study of the Wenzhou "Financial Brain" platform* (Master's thesis, Dongbei University of Finance and Economics). https://doi.org/10.27006/d.cnki.gdbcu.2024.001055

[7] Mao, Z., & Liu, J. (2021). A review of credit risk assessment based on machine learning methods. *Economic Research Guide*, (23), 117–119.

[8] Heß, V. L., & Damásio, B. (2025). Machine learning in banking risk management: Mapping a decade of evolution. *International Journal of Information Management Data Insights, 5*(1), 100324. https://doi.org/10.1016/j.jjimei.2025.100324

[9] Tian, X., Tian, Z., Khatib, S. F. A., & Wang, Y. (2024). Machine learning in internet financial risk management: A systematic literature review. *PLOS ONE, 19*(4), e0300195. https://doi.org/10.1371/journal.pone.0300195

[10] Ding, S., & Wang, H. (2026). Artificial intelligence and financial security: Technical vulnerabilities, risk evolution paths, and prevention mechanisms. *Reform of Economic System*, (2).

[11] Mo, W., & Xu, R. (2025). Risks and responses of artificial intelligence applications in the financial sector. *New Finance*, (11), 24–33.

[12] Zhang, H., Li, Z., Jing, Y., Wang, X., Wu, J., & Chai, H. (n.d.). Research on financial risk monitoring and early warning system under the background of digital transformation. *Strategic Study of CAE*. https://doi.org/10.15302/J-SSCAE-2024.03.020