Proceedings of ICMRED 2026 Symposium: The Future of Work: Strategy, Workforce Transformation, and Organizational Renewa
DOI: 10.54254/2754-1169/2026.LD34882

A Multidimensional Comparison of Artificial Intelligence Methods Across Different Financial Risk Control Scenarios

Yueran Yang

Harbin Institute of Technology, Harbin, China
2025112930@stu.hit.edu.cn

Abstract. The four core categories of financial risk—credit risk, market risk, operational risk, and liquidity risk—differ significantly in data characteristics and modeling target. Thus, traditional statistical methods and conventional machine learning models often fail to meet practical demands. In recent years, deep learning, such as graph neural networks and Bayesian methods, has offered new approaches. However, most existing researches focus on a single type of risk or a specific model architecture, and systematic comparisons of model applicability across different scenarios remain limited. This paper starts from the four core types of financial risk, comparing and analyzing more than ten models from dimensions, including predictive performance, interpretability, data dependence, computational cost, regulatory compatibility and so on. The study demonstrates that different methods have their own advantages under different scenarios. Obvious discrepancies exist in interpretability, extreme-risk characterization, and regulatory compliance. Furthermore, the paper discusses the future development of intelligent risk control systems from the perspectives of model integration, dynamic adaptation, and regulatory technology.

Keywords: Financial risk management, artificial intelligence, model comparison, deep learning

1. Introduction

Financial risk management is the foundation of the stable operation of financial institutions. Traditional statistical methods and conventional machine learning models struggle to address complex data characteristics and cater to market demands. Whereas deep learning, graph neural networks, Bayesian methods, and their hybrid models have obvious advantages in nonlinear fitting, structured modeling, and uncertainty representation. These methods are gradually used in the field of risk control to improve predictive performance. Most existing studies focus on a single type of risk or a specific model. A systematic review of model applicability in different scenarios remains insufficient. In addition, the limitations of these models and their compatibility with regulatory requirements still require further attention. Thus, this paper cuts into four core types of financial risk--credit risk, market risk, operational risk, and liquidity risk--to review their current applications and model characteristics, and to compare the strengths and limitations of different technical approaches. It also evaluates the merits of multiple methods and summarizes future development

DOI: 10.54254/2754-1169/2026.LD34882

directions to provide a useful reference for the construction and improvement of intelligent risk control systems.

2. Credit risk

Credit risk mainly arises from counterparty default. Its data are characterized by severe class imbalance, the low frequency of default events, and strict regulatory requirements for interpretability. These features place higher demands on model discrimination accuracy, imbalanced learning capability, and decision transparency. This paper takes the empirical study of Hu Mingyuan as an example for analysis [1].

2.1. The dual-path design of the E-MLP and B-MLP models

To satisfy the requirements of computational cost, deployment speed, and statistical rigor in different business scenarios, study [1] adopts the following two technical paths.

E-MLP deep ensemble model: This model integrates multiple MLP subnetworks and uses the ensemble mean of the empirical distribution as a calibrated point estimate to correct calibration bias. It implicitly quantifies epistemic uncertainty through output differences among subnetworks. Logit-Normal parameters explicitly model aleatoric uncertainty in the data. Quantiles are extracted to construct confidence intervals, enabling flexible threshold setting and differentiated pricing. This model is suitable for online or near-online risk control scenarios with large-scale data.

B-MLP deep Bayesian model: This model uses Gaussian priors and variational inference to approximate the posterior distribution of parameters, capturing the coupled effects of data noise and model uncertainty. It applies Monte Carlo sampling to obtain the posterior distribution of predicted probabilities and the highest posterior density interval, which provides strong interpretability and statistical rigor. Prior regularization reduces the risk of overfitting, making this model suitable for small-data settings or scenarios with high interpretability requirements.

2.2. Limitations and future prospects of the E-MLP and B-MLP dual-path models

The probability calibration of E-MLP relies on post hoc calibration methods, which makes it difficult to correct extreme-risk intervals. In addition, the model lacks an explicit probabilistic framework, so its uncertainty estimates may be less reliable. B-MLP, by contrast, is constrained by strong parameter assumptions and cannot capture certain feature patterns effectively. Its high computational cost also limits its applicability in large-scale data environments or real-time risk control scenarios. More flexible prior distributions are therefore needed to improve their suitability for financial applications.

A common future direction lies in the development of uncertainty-aware and interpretable methods that better satisfy regulatory requirements, while also improving inference algorithms. In addition, probabilistic intervals should be used to design dynamic risk thresholds, so as to achieve adaptive risk control.

2.3. Result analysis of the E-MLP and B-MLP dual-path models

As shown in Table 1, the two models both perform well across three datasets. In terms of discriminatory power, the AUC ranges from 0.711 to 0.762, representing an improvement of up to 12.8% over traditional models. As for calibration performance, Beta calibration reduces the Expected Calibration Error (ECE) by more than 74%, while the Brier Score (BS) reaches a

DOI: 10.54254/2754-1169/2026.LD34882

minimum of 0.143. In addition, both models provide interval estimates of default probability and quantify predictive uncertainty, balancing accuracy and interpretability.

Table 1. Comparison of models in the credit risk case

Dimension	Traditional Methods	E-MLP	B-MLP
Predictive performance	Statistical models, tree-based models, and conventional MLPs show stable performance	Outperforms traditional models in both discrimination and calibration	Slightly outperforms E-MLP in discrimination and shows strong generalization ability
Interpretability	Statistical models are generally more interpretable than deep learning models	Introduces a feature-weighting network and SHAP-based attribution to improve the interpretation of multiple risk factors and their interactions	Provides a Bayesian probabilistic explanation pathway, with more transparent feature contributions and stronger interpretability
Data requirements	Traditional models are sensitive to data volume	Suitable for large-scale datasets	Suitable for small-sample scenarios
Computational cost	Low	Efficient due to parallel training, relatively low	Variational inference is complex and computationally expensive
Real-time capability	Strong	Relatively strong and suitable for large-scale data scenarios	Weak, heavy computational demands
Regulatory friendliness	Statistical models are easier to align with regulatory requirements	Offers confidence intervals and uncertainty quantification, with good regulatory compatibility	Statistically rigorous, with stricter confidence intervals and stronger regulatory acceptability

3. Market risk

Market risk mainly arises from fluctuations and the network-based transmission of asset prices. Its date has characteristics such as high frequency and continuity, temporal dependence, volatility clustering, and nonlinear associations, which place higher demands on a model's ability to capture temporal dynamics and structural relationships. This paper takes the empirical study of B. Padmavathi and P. Ramanathan as an example for analysis [2].

3.1. Conceptual design of a GNN-based model for volatile market risk

This model introduces a graph neural network (GNN)-based technical approach. Financial entities are treated as nodes, and the complex relationships among them are represented as edges, constructing a dynamic financial graph. It also integrates techniques such as GCN, TGN, and GAT to capture both the spatial contagion of risk and its temporal evolution. Finally, it simulates the diffusion process of risk and outputs risk assessment results, volatility forecasts, and rankings of risky entities, which can be applied to risk control practices such as stress testing and anomaly detection. Figure 1 lists the flowchart of the volatility market risk model based on graph neural networks.

DOI: 10.54254/2754-1169/2026.LD34882

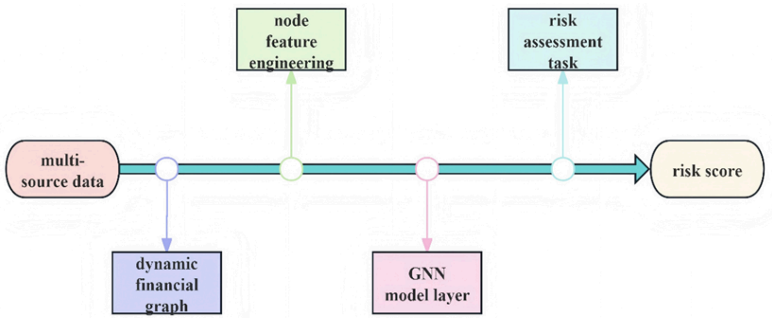

Figure 1. Flowchart of the volatility market risk model based on graph neural networks

3.2. Analysis of results from the GNN-based volatility market risk model

In the task of volatility forecasting, the proposed model achieved a root mean square error (RMSE) of 0.059, substantially outperforming the LSTM model (0.080) and the random forest model (0.088). For the identification of the top ten high-risk entities, the model attained a Precision@10 of 0.88, representing an improvement of over 25% compared with conventional approaches. Furthermore, during the COVID-19 period, the energy sector (0.91) and the financial sector (0.85) were identified as the highest-risk industries, a finding consistent with actual market performance and validating the model's practical relevance, as shown in Table 2.

Table 2. Comparison of models in the market risk case

Dimension	Traditional Methods (Logistic Regression, Random Forest, LSTM, etc.)	GNN-based Methods (including GCN and GAT)
Predictive performance	Moderate accuracy, with relatively limited overall performance; F1-score and RMSE are generally inferior	Clearly improve predictive accuracy, with better F1-score and RMSE
Interpretability	Partially interpretable	GAT improves interpretability through the attention mechanism, enabling the identification of key nodes and relationships and revealing risk transmission paths
Data requirements	Mainly rely on time-series data and individual features, with limited use of structural relationships	Require the construction of a time-evolving financial graph and the integration of multi-source data, including historical prices, macroeconomic indicators, industry relationships, and news sentiment
Computational cost	Relatively low computational burden and faster model training	Higher computational complexity and longer training time
Temporal modeling capability	Have certain limitations in handling complex dependencies	Stronger capability, as graph structures effectively capture complex dependencies among nodes

3.3. Limitations and future prospects of the GNN-based model for volatile market risk

This model involves high computational and training costs, and the required data are often difficult to obtain and process. Future work should integrate multi-source heterogeneous data in order to capture entity relationships and market sentiment more comprehensively. Its interpretability also remains insufficient, which calls for the development of interpretable GNN methods that better satisfy regulatory requirements. In addition, existing approaches are often limited to a single graph

DOI: 10.54254/2754-1169/2026.LD34882

structure. Dynamic graph learning should therefore be incorporated to improve adaptability and timeliness in response to changing market conditions.

4. Operational risk

In this paper, fraud risk is taken as a representative example of operational risk. Fraud risk mainly arises from malicious transactions. Fraud samples are infrequent and discrete, and fraud patterns evolve continuously over time. These characteristics place high demands on a model's ability to detect anomalies and to learn adaptively. This paper takes the empirical study of B. Padmavathi and P. Ramanathan as an example for analysis [3].

4.1. Result analysis of the CNN–LSTM two-stage architecture

As shown in Table 3, ogundokun et al. report an accuracy of 99.72% in Ethereum phishing detection. Asiri et al., using a graph convolutional network, achieve an accuracy of 98.56% and an AUC of 0.9444 in Bitcoin transaction analysis. Building on these studies, the SecureChainNet model proposed by Padmavathi et al. significantly outperforms traditional methods in terms of accuracy, precision, recall, and related metrics. Its area under the ROC curve approaches 1.0, indicating that it can effectively identify complex fraud patterns while also offering strong interpretability.

Table 3. Comparison of models in the fraud risk case

Dimension	SecureChainNet	Traditional Methods (CNN, LSTM, Random Forest, XGBoost, etc.)
Predictive performance	Outperforms baseline models in accuracy, F1-score, precision, and recall; achieves a high AUC and strong overall classification performance	Lower performance, with a higher tendency toward false positives and false negatives
Interpretability	Supports modular transparency; the CNN stage generates heat maps of high-risk features, while the LSTM stage visualizes attention weights, improving interpretability	Relatively weak, as most methods function as black-box models
Data requirements	Uses both raw and engineered features; supports federated learning, protects privacy, and does not require centralized raw data	Some methods rely on centralized data and require a relatively large amount of labeled data
Capability for detecting fraudulent groups	Strong, as it uses graph-theoretic features and wallet behavior sequences to help identify complex fraud-group patterns	Weak, as many methods ignore complex spatiotemporal dependencies

4.2. Conceptual design of the CNN–LSTM two-stage architecture

To address the data characteristics, this paper designs a CNN–LSTM two-stage architecture. The CNN layer identifies anomalous transactions in the feature space, while the LSTM layer detects fraud strategies with temporal regularities, such as money-laundering cycles, to meet the demands of real-time and accurate fraud detection in risk-sensitive financial scenarios. In addition, the interpretability of the model is enhanced through optimizations such as the integration of attention mechanisms.

DOI: 10.54254/2754-1169/2026.LD34882

4.3. Limitations and future prospects of the CNN–LSTM two-stage architecture

The model structure is relatively complex, computationally expensive, and highly dependent on data quality and availability. Future improvements may incorporate graph neural networks and multimodal data to strengthen feature extraction. Online learning mechanisms can also be introduced to better support edge-based real-time response scenarios. In addition, the combination of federated learning and explainable artificial intelligence can help protect data privacy while meeting regulatory requirements.

5. Liquidity risk

Table 4 lists the comparison of models in the liquidity risk case. Liquidity risk mainly refers to the inability of a firm to meet its debt obligations at a reasonable cost. It is characterized by strong contagion effects, significant sensitivity to market sentiment, and pronounced temporal dependence. Effective modeling therefore, relies on real-time monitoring capability and the ability to capture nonlinear relationships. This paper takes the model proposed by Agarwal and Agrawal as an example for analysis [4].

5.1. Conceptual design of CapitalFlowNet

The model mainly consists of three components. The first is an LSTM-based temporal forecasting network, which captures cyclical fluctuations in cash flow from historical transaction data in order to predict liquidity demand. The second is a supplier risk classification module, which uses a deep neural network to generate dynamic risk scores based on behavioral data. The third is a capital allocation optimizer, which dynamically adjusts credit limits under risk scores and budget constraints through linear optimization, controlling overall risk while prioritizing suppliers' liquidity needs.

5.2. Result analysis of CapitalFlowNet

The model achieves a 91.2% accuracy rate for liquidity demand, a capital utilization efficiency of 88.6%, and an anomaly response rate as high as 93%, significantly outperforming traditional heuristic methods and single machine learning models. It demonstrates strong feasibility in real-time liquidity monitoring, dynamic capital allocation, and liquidity risk early warning.

5.3. Limitations and future prospects of CapitalFlowNet

The model depends heavily on the quality of historical data, which suggests the need to incorporate blockchain-based and multimodal data sources. It also suffers from high computational cost, limited interpretability, delayed responses to sudden events, insufficient modeling of external risks, and a relatively narrow evaluation framework. In addition, its long-term stability in practical applications remains underexplored. Future research should establish online learning mechanisms to improve real-time responsiveness and adopt multilevel supply-chain coordination strategies.

DOI: 10.54254/2754-1169/2026.LD34882

Table 4. Comparison of models in the liquidity risk case

Dimension	CapitalFlowNet	LSTM	Deep Reinforcement Learning	Heuristic Methods
Predictive capability	Strong	Relatively strong	Relatively weak	Weak
Real-time capability	Strong.	Relatively weak	Relatively weak	Weak
Interpretability	Equipped with a decision dashboard that supports visualization and monitoring	Mainly black-box models with limited interpretability	Complex models with relatively weak interpretability	Rule-based and relatively interpretable
Macroeconomic sensitivity	High	Relatively high	Relatively low	Low

6. Discussion

To systematically compare the characteristics of different models, this paper summarizes their performance according to the study of credit risk [5, 6], market risk [7], fraud risk [8, 9] and liquidity risk [10] across key dimensions, as shown in Table 5.

Table 5. Comparison of commonly used models in financial risk analysis

Dimension	Logistic Regression	Random Forest	Isolation Forest	XGBoost	SVM	CNN	GNN	LSTM/GRU	Transform
Predictive performance	Moderate, as a linear baseline.	High	Moderate (for anomaly detection)	High	Moderate	High (for image/sequence tasks)	High (for capturing relationships)	High	High
Interpretability	High	Moderate	Low	Moderate	Low, especially with nonlinear kernels	Low	Low	Low	Low
Data requirements	Low	Relatively low, with lower risk of overfitting	Low	Moderate, with attention needed to avoid overfitting	Moderate	High, especially for large datasets	Moderate	High	High
Real-time capability	High	Relatively high	Relatively high	Relatively high	Relatively high	Moderate	Moderate	Moderate	Low
Ability to handle imbalanced data	Low	High	High	High	Low	Moderate	Moderate	Moderate	Moderate

DOI: 10.54254/2754-1169/2026.LD34882

Table 5. (continued)

Regulatory friendliness	High	Moderate	Low	Moderate	Low	Low	Low	Low	Low
Model-specific ability	Outputs probabilities with relatively clear calibration meaning	Resistant to overfitting; supports feature selection	Unsupervised anomaly detection	Regularization helps prevent overfitting; handles missing values	Effective for high-dimensional data	Local feature extraction	Learns node representations in graph-structured contexts	Captures temporal dependencies through gating mechanisms	Parallel processing and global attention

7. Conclusion

This paper, through a multidimensional comparative analysis of artificial intelligence models across four core categories of financial risk—credit risk, market risk, operational risk, and liquidity risk—finds that different models possess significant advantages in their respective application scenarios. E-MLP and B-MLP balance predictive accuracy and uncertainty quantification in credit risk assessment. GNN-based models effectively capture risk contagion paths in market risk analysis. CNN–LSTM architectures accurately identify spatiotemporal anomalies in fraud detection. And CapitalFlowNet improves the efficiency of liquidity management through a prediction–optimization framework. However, one single model can not fully satisfy the business demands of diversity, dynamic change, and regulatory compliance. Each model involves its own advantages and disadvantages among interpretability, extreme-risk characterization, and computational cost, as summarized in Table 5.

To promote the accurate, transparent, and efficient development of risk control investment in the future, further studies should mainly pay attention to the following directions. First, based on the differentiated requirements of specific scenarios for accuracy, timeliness, transparency, and compliance, hybrid modeling strategies should be constructed. For example, in credit risk assessment, the integration of Transformer and XGBoost can simultaneously capture sequential transaction patterns and static information, achieving complementary modeling of dynamic behavior and structured data. In fraud detection, the combination of Transformer and graph neural networks can jointly account for transaction-sequence dependence and the structure of capital flow networks, improving the identification of organized fraud. By integrating the strengths of different models, such hybrid approaches can better match the demands of real-world scenarios. Second, dynamic adaptability should be strengthened. The use of online learning and the incorporation of multimodal data can improve a model's ability to handle extreme events and enhance its responsiveness to practical problems promptly. Third, regulatory compatibility should be reinforced by improving model interpretability and ensuring thatat regulatory requirements are met while protecting the privacy of personal data.

References

[1] Hu, M. Y. (2025). Research on credit risk assessment based on Bayesian deep learning [Master's thesis]. Shandong University of Finance and Economics.

DOI: 10.54254/2754-1169/2026.LD34882

[2] Narsepalle, K. R., & Bolla, S. R. (2025). Graph neural networks for advanced financial risk modeling in volatile markets. In 2025 International Conference on Engineering, Technology & Management (ICETM) (pp. 1–5). IEEE.

[3] Padmavathi, B., & Ramanathan, P. (2025). SecureChainNet: A dual-stage CNN-LSTM model for real-time transaction fraud detection in cryptocurrency networks. In 2025 International Conference on Modern Sustainable Systems (CMSS) (pp. 734–742). IEEE.

[4] Agarwal, S., & Agrawal, J. (2025). AI-powered forecasting models for optimizing working capital in supply chain financing. In 2025 World Skills Conference on Universal Data Analytics and Sciences (WorldSUAS) (pp. 1–6). IEEE.

[5] Raliphada, P., Olusanya, M. O., & Olukanmi, S. (2025). Optimizing credit risk model classification using BERT, RoBERTa, and LLaMA 3.2. In IEEE EUROCON 2025 - 21st International Conference on Smart Technologies (pp. 1–6). IEEE.

[6] Li, Q. (2023). Research on bank credit risk assessment based on BP neural network. In 2023 2nd International Conference on 3D Immersion, Interaction and Multi-sensory Experiences (ICDIIME) (pp. 322–326). IEEE.

[7] Tang, D. (2023). Optimization of financial market forecasting model based on machine learning algorithm. In 2023 International Conference on Networking, Informatics and Computing (ICNETIC) (pp. 473–476). IEEE.

[8] Beri, M., Gill, K. S., & Sharma, N. (2024). Enhancing credit card fraud detection: A comparative analysis of machine learning models. In 2024 4th International Conference on Sustainable Expert Systems (ICSES) (pp. 449–454). IEEE.

[9] Owen, B., Tanto, M. J., Law, S., & Agustia Rachmawati, I. D. (2025). Implementing isolation forest anomaly detection into a hybrid K-Means & K-NN algorithm for credit card fraud detection. In 2025 13th International Conference on Awareness Science and Technology (iCAST) (pp. 1–5). IEEE.

[10] Paramarta, V., Tampubolon, L. P. D., Rahman, A., Safitri, A. N. I., Mardiyah, A. M., & Dewi, I. K. (2025). Comparative analysis of SVM, decision tree, and XGBoost for corporate liquidity risk classification based on Indonesian stock exchange financial data. In 2025 Tenth International Conference on Informatics and Computing (ICIC) (pp. 1–6). IEEE.