ISSN 2616-5902 Vol. 7, Issue 4: 98-103, DOI: 10.25236/AJBM.2025.070412

Based on Machine Learning: Research on Improving the Liquidity Risk Identification Model of Commercial Banks

Qianyi Chen[1,a], Xin Song[2,b,*]

[1]*Business School, University of Shanghai for Science and Technology, Shanghai, China*
[2]*Business School, University of Shanghai for Science and Technology, Shanghai, China*
[a]*qianyi_279@163.com*, [b]*songxin@usst.edu.cn*
**Corresponding author*

Abstract: *The changing economic environment is fraught with uncertainty, making the prevention and resolution of financial risks and the enhancement of liquidity management capabilities key focuses and challenges for financial institutions and regulatory authorities in China. With the support of artificial intelligence and big data technologies, the research paradigms for liquidity risk management have been continuously enriched. Machine learning models have been proven to possess unique advantages in the field of liquidity risk identification. This paper summarizes the main application frameworks of machine learning in current risk management research, along with their applicability and limitations. It further explores potential future improvements and research trends in applying machine learning to identify liquidity risks.*

Keywords: *Machine learning; Risk identification; Liquidity risk; Commercial banks*

1. Introduction

In recent years, Chinese commercial banks have strengthened the application of financial technology and advanced digital transformation to cope with the rapidly changing financial market environment, focusing on preventing and resolving financial risks. Relevant liquidity risk regulatory policies have increasingly aligned with the latest changes in the industry[1]. However, in the long term, external financial market uncertainties and macroeconomic fluctuations, among other factors, may still pose new challenges to the liquidity risk management of commercial banks. Strengthening the liquidity management capabilities of China's banking sector and all financial institutions remains a critical focus and challenge.

With the support of artificial intelligence and big data technologies, scholars have applied new technologies such as natural language processing and computer vision, as well as new methods like deep learning and machine learning, to enrich the research paradigms of liquidity risk management. Many studies have adopted machine learning methods to explore topics such as systemic risk identification in banks[2], risk identification in online lending platforms[3], and liquidity risk in credit bonds[4]. However, there are few application cases specifically addressing liquidity risk management in banks. Compared to traditional empirical models, machine learning is not constrained by general forms of variable relationships and allows for the existence of interactive and multivariate relationships. It can also consider high-dimensional control variables, enabling more accurate estimation of causal relationships[5]. Additionally, in terms of model prediction effectiveness, machine learning can effectively capture nonlinear interactions between variables and overcome the limitations of linear regression in cases where variables are highly correlated[4]. Nevertheless, the application of machine learning models in the field of risk management remains relatively narrow, and research on liquidity risk management in commercial banks is limited. By constructing machine learning models to identify bank liquidity risks, it is possible to analyze the sources of liquidity risks in a timely and accurate manner, allowing regulatory authorities and banks to take prompt action and implement effective measures to address potential crises caused by liquidity risks. In summary, introducing machine learning models into the field of liquidity risk management to break through existing research paradigms has significant practical implications.

This study focuses on the application of machine learning in the identification of liquidity risks in

ISSN 2616-5902 Vol. 7, Issue 4: 98-103, DOI: 10.25236/AJBM.2025.070412

commercial banks. Based on existing literature, it analyzes the applicability and limitations of machine learning models, proposes theoretical improvements, and further explores future directions for applying machine learning in liquidity risk identification.

2. Application of Machine Learning in Risk Management

2.1 Concepts Related to Machine Learning

With the support of artificial intelligence and big data technologies, scholars have applied new technologies such as natural language processing and computer vision, as well as new methods like deep learning and machine learning, to enrich the research paradigms of risk management. Compared to traditional empirical models, machine learning is not constrained by general forms of variable relationships and allows for the existence of interactive and multivariate relationships. It can also consider high-dimensional control variables, enabling more accurate estimation of causal relationships[5]. Additionally, in terms of model prediction effectiveness, machine learning can effectively capture nonlinear interactions between variables and overcome the limitations of linear regression in cases where variables are highly correlated[4]. The general process of machine learning is illustrated in Figure 1.

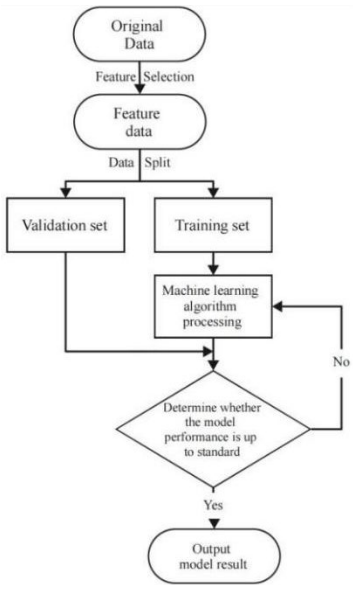

Figure 1. Machine learning flow chart

2.2 Specific Applications in Risk Management

Machine learning models have been proven to outperform traditional factor identification and early warning analysis models in various financial fields. Zhu et al. (2016) applied random forest (RF), random subspace (RS), random AdaBoost (RAB), and an ensemble of RS and RAB (RS-RAB) to predict credit risks for small and medium-sized enterprises (SMEs) in China's supply chain finance. The results showed that RF had the fastest learning speed and the highest accuracy, successfully avoiding overfitting[8]. In the study of P2P lending platform risk identification, Cui Yanyan and Liu Lixin (2023) used logistic regression and random forest models to evaluate the importance of indicators in the risk assessment system. The results indicated that SVM and random forest models performed better on the test set[9]. Li Shuang et al. (2022) scientifically classified platform risk levels and tested the early warning performance of naive Bayes (NB), Softmax regression, random forest (RF), and multi-class support vector machines (SVMs)[3]. Zhang Zongxin and Zhou Cong (2024) used regression tree models such as random forest (RF) and extreme gradient boosting regression trees (XGBoost), as well as neural network models, to measure and warn of liquidity risks in credit bonds. They concluded that neural network models with one hidden layer had the strongest early warning capabilities[4]. Shi Rong et al. (2024) reviewed a large body of literature, analyzing the application value of machine learning in economic forecasting and summarizing the applicability of different machine learning algorithms in financial and economic research. These studies provide valuable references for other commercial risk studies, including liquidity risk[10].

Research on machine learning models for commercial bank risk is more diverse, but most studies are still limited to the application of single models, leaving room for improvement in model performance. Tavana et al. (2018) used fuzzy inference and neural network techniques to provide early warnings for commercial bank risks. The results showed that fuzzy neural networks (FNN) could effectively classify and predict risks[11]. To identify systemic risks in banks, Li Chenying (2020) proposed a systemic risk identification model that included local outlier factor (LOF), isolation forest (IF), and angle-based outlier detection (ABOD) algorithms[2]. Triepels et al. (2021) applied recurrent neural networks (RNN) to monitor bank liquidity management by detecting anomalies in large-value payment system (LVPS) transaction logs. Using multivariate Gaussian classifiers and different RNN variants, they demonstrated that anomaly detection could be an effective tool for bank regulators to identify irregular behaviors[12]. Other studies have used backpropagation neural networks (BPNN) for early warning and assessment of commercial bank risks[13], or employed LightGBM to classify liquidity risks and empirically compare the robustness of various models, such as logistic regression, support vector machines, and XGBoost, in testing bank liquidity risks[14]. Additionally, perceptron neural networks have been developed to predict local bank liquidity risks[15]. The summary of machine learning research in risk management be showed in Table 1.

Table 1: Summary of Machine Learning Research in Risk Management

Machine Learning Model	Specific Classification	Application Scholars
XGBoost	Ensemble Learning	Tan and Gan (2022); Zhang and Zhou (2024)
RF	Ensemble Learning	Dai and Shen (2022)[6]; Zhu, et al (2016); Cui and Liu (2020); Lee, et al (2022); Zhang and Zhou (2024)
RS	Ensemble Learning	Zhu, et al (2016)
RAB	Ensemble Learning	Zhu, et al (2016)
RNN	Deep Learning	Triepels, et al (2021)
MLP	Deep Learning	Lu and Zhang (2022)[7]; Sumi (2024); Zhang and Zhou (2024)
BPNN	Deep Learning	Yang, et al (2021)
FNN	Deep Learning	Tavana, et al (2018)
IF	Anomaly Detection	Chenying Lee (2020)
LOF	Anomaly Detection	Chenying Lee (2020)
ABOD	Anomaly Detection	Chenying Lee (2020)
PCA	Dimensionality Reduction	Zhang and Zhou (2024)
Logistic	Classification Model	Tan and Gan (2022); Dai and Shen (2022); Lee, et al (2022); Cui and Liu (2020)
SVM	Classification Model	Tan and Gan (2022); Cui and Liu (2020)
SVMs	Classification Model	Lee, et al (2022)
NB	Classification Model	Lee, et al (2022); Cui and Liu (2020)
LightGBM	Classification Model	Tan and Gan (2022)

2.3 Summary of Machine Learning Applications in Risk Management

Existing literature has comprehensively and deeply explored the issue of liquidity risk in commercial banks from various perspectives and has accumulated rich research results. However, there is still room for improvement. For example, in practice, liquidity risk management involves the entire process of identifying, measuring, evaluating, and controlling liquidity risks, but existing literature often focuses on only one aspect of this process. Additionally, research on machine learning in risk management is relatively limited, and model applications are relatively narrow.

Based on the current state of machine learning applications in risk management, it is evident that in the identification of liquidity risks, machine learning breaks through traditional limitations by no longer relying solely on financial indicators and a few influencing factors. For instance, exploring the impact of economic policy uncertainty and other factors, machine learning can mine risk signals from multiple factors, process large-scale data with complex algorithms, and accurately capture risk correlations, providing a comprehensive perspective for risk identification and enhancing the accuracy and timeliness of identification. In terms of risk management tools, machine learning complements traditional methods by integrating internationally recognized indicators and advanced internal pricing tools for monitoring and early warning, enabling real-time tracking of risk dynamics and optimizing resource allocation to improve management efficiency. In various fields of risk identification and management, machine learning models have demonstrated excellent performance. For example, random forests have proven effective in predicting credit risks for SMEs in supply chain finance; SVM and random forest models have shown reliable early warning capabilities in P2P lending credit risks; and neural network models have exhibited strong early warning capabilities in the credit bond market.

Although machine learning has been proven to have higher accuracy and stronger applicability in many fields, its application in the field of liquidity risk management in banks is not yet widespread, with most studies relying on single models. The identification and analysis of liquidity risks in commercial banks is a complex and important topic. In the future, it is necessary to deepen the understanding of the concept of liquidity risk, effectively identify liquidity risks, and adopt scientific management methods and tools. At the same time, more reasonable machine learning models should be combined to effectively identify and control risks, enabling Chinese commercial banks to better cope with liquidity risk challenges and ensure their stable operation and sustainable development.

3. Improvement of Commercial Bank Liquidity Risk Identification Models Based on Machine Learning

Based on a review of the literature on risk management, this paper suggests that future research on applying machine learning to liquidity risk identification could focus on combining multiple machine learning models.

From the analysis of deep learning and ensemble learning models, it is evident that deep learning and ensemble learning are complementary in some ways. Deep learning can automatically learn feature representations through neural networks but may be prone to overfitting. Ensemble learning, by combining multiple base learners, can improve the generalization ability and accuracy of models but may require more computational resources. Therefore, combining deep learning and ensemble learning can achieve more powerful models.

If the research focus is on liquidity risk management in listed commercial banks in China, the data in this context typically has complex multi-dimensional characteristics and may involve nonlinear relationships. Based on a comprehensive evaluation of the above models, random forest (RF) and multilayer perceptron (MLP) are more suitable for this research task. RF can handle various types of features, especially the complex feature relationships in bank data, and can evaluate key variables through feature importance, which is useful for identifying key factors in liquidity risk analysis. MLP, as the foundation of deep learning, has strong nonlinear modeling capabilities and can better capture potential complex patterns in bank liquidity risks. These two models have significant advantages in comprehensively analyzing, predicting, and managing bank liquidity risks. This paper selects RF from ensemble models and MLP from deep learning models as examples to analyze future directions for combining multiple models.

Based on the structural characteristics of single machine learning models and the summarized advantages and disadvantages of each model, this paper proposes five improvement ideas to enhance the performance of hybrid models by optimizing feature selection, improving classification results, and other aspects.

Cascading Models: Use RF and MLP as base models and combine them through stacking. First, use RF and MLP to make preliminary predictions, and then use the prediction results (probability outputs or category labels) as new input features. Finally, use a simple model, such as linear regression, as a meta-learner to make the final liquidity risk identification rating. This approach leverages the feature selection capabilities of RF and the nonlinear modeling capabilities of MLP to improve the overall model's accuracy and robustness.

Feature Fusion: Combine the feature extraction capabilities of the two models. In this method, RF and MLP can be used to extract features at different levels. Specifically, use RF to perform feature selection or dimensionality reduction on the original input data and output important features. Then, input the important features into MLP for further deep learning. This method allows RF to filter out the most critical features for liquidity risk identification, while MLP is responsible for in-depth modeling of the refined features, thereby improving prediction performance.

Joint Training: RF and MLP models work together through joint training. First, train the RF model

ISSN 2616-5902 Vol. 7, Issue 4: 98-103, DOI: 10.25236/AJBM.2025.070412

to obtain preliminary prediction results and analyze feature importance to obtain initial feature weights. Then, use the feature importance output by RF as the initial weights for the MLP model to further optimize the model, allowing MLP to continue learning and adjusting based on the preliminary results. This approach enables MLP to effectively utilize the feature importance information from RF, accelerating model convergence and enhancing the understanding of complex data.

Weighted Model Fusion: Directly weight and average the output results of RF and MLP models. Specifically, train RF and MLP separately to obtain their prediction probabilities. Then, combine the prediction results through weighted averaging to generate the final classification probability. The weighting coefficients can be dynamically adjusted based on cross-validation results to ensure the model's adaptability in different scenarios. This method is simple and feasible, and by assigning appropriate weights to each model, the combined model's performance can be optimized.

Soft Voting: Make decisions based on the probability outputs of multiple models. Specifically, let RF and MLP output the probability values for each class. Then, weight and average the probability values of the two models and select the class with the highest weighted probability as the final identification result. Soft voting can fully utilize the different prediction characteristics of the two models, avoiding biases that may arise from a single model and improving the stability of identification.

RF excels in handling high-dimensional features and noisy data, effectively selecting important features. MLP is suitable for modeling complex nonlinear relationships, especially for highly nonlinear data. By combining these two models, their strengths can be integrated. In summary, a hybrid model based on RF and MLP can more comprehensively capture the complexity of bank liquidity risks, improving the accuracy and robustness of liquidity risk identification. The structure of the hybrid model is shown in Figure 2.

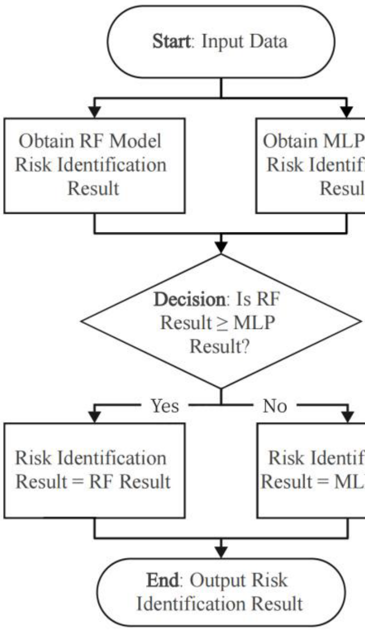

Figure 2. Machine learning hybrid model flow chart

4. Conclusion

With the rapid development of information technology, big data and machine learning have brought new opportunities for research on liquidity risk identification in commercial banks. This paper systematically reviews the development of machine learning and risk identification research, and deeply explores the application methods of their integration. This helps to clarify the advantages of machine learning in terms of data processing capabilities and model construction flexibility, as well as its limitations in areas such as result interpretation and data requirements. On this basis, further exploration of optimized application methods for machine learning can enable it to play a greater role in research related to commercial bank liquidity risks, providing strong support for the stable operation of commercial banks.

ISSN 2616-5902 Vol. 7, Issue 4: 98-103, DOI: 10.25236/AJBM.2025.070412

By summarizing existing research findings, the advantages of machine learning in the field of liquidity risk identification are significant. In terms of data utilization, machine learning enables the acquisition of large-scale data, far exceeding the data volume under traditional research paradigms, providing a solid foundation for precise analysis. It also broadens data sources, no longer limited to traditional information channels such as corporate annual reports and financial statements, but extending to news, media, and other levels, comprehensively capturing relevant information. Additionally, machine learning encompasses diverse data formats, including structured, semi-structured, and a large amount of unstructured data. Compared to traditional methods, the powerful data processing capabilities of machine learning allow it to effortlessly analyze massive and complex data, uncovering hidden patterns. As a result, liquidity risk identification research has broken through traditional limitations. By fully integrating various market information and deeply exploring variables closely related to market changes that are beyond the scope of traditional research, machine learning ultimately achieves more accurate risk assessment results. Furthermore, in the face of the overwhelming flood of complex data, how to efficiently filter out truly valuable and effective data poses a new challenge for researchers, necessitating innovative solutions.

In summary, machine learning is revolutionizing the research paradigms and methods in the field of risk management. Its application in liquidity risk identification research still has room for improvement and is worthy of in-depth exploration by scholars.

References

[1] Dong Xiaolin, Wu Zhiwei, Chen Qiuyue. The Impact of Financial Technology Development on Risk Prevention and Control in Commercial Banks—An Empirical Analysis Based on 176 Commercial Banks in China [J]. Jiangsu Social Sciences, 2023, (01): 84-94+242-243.(In Chinese)

[2] Lee Chenying. Research on the Identification and Early Warning of Systemic Risks in China's Commercial Banks [J]. Journal of Central University of Finance & Economics, 2020, (10): 36-53.(In Chinese)

[3] Lee Shuang, Liu Yaya, Lee Xiaotong. Risk Monitoring and Early Warning of P2P Lending Platforms Based on Machine Learning [J]. Journal of Xi'an University of Finance and Economics, 2022, 35(02): 62-73.(In Chinese)

[4] Zhang Zongxin, Zhou Cong. Research on Early Warning of Credit Bond Liquidity Risk Based on Machine Learning [J]. International Finance Research, 2024, (05): 74-85.(In Chinese)

[5] Ni Xuanming, Zheng Tiantian, Zhao Huimin. Can Civil-Military Integration Reduce the Equity Financing Costs of High-Tech Enterprises?—An Empirical Study Based on Double Machine Learning [J]. Systems Engineering—Theory & Practice, 2023, 43(06): 1630-1650.(In Chinese)

[6] Dai Yarong, Shen Yifeng. Can Random Forest Models Predict Bond Defaults in China? [J]. Journal of Econometrics, 2022, 2(02): 418-440.(In Chinese)

[7] Lu Zhe, Zhang Jian. Credit Risk Prediction of Listed Companies Based on SMOTETomek-RFE-MLP Algorithm [J]. Systems Science and Mathematics, 2022, 42(10): 2712-2726.(In Chinese)

[8] Zhu Y, Sun B, et al. Predicting China's SME Credit Risk in Supply Chain Financing by Logistic Regression, Artificial Neural Network and Hybrid Models[J]. Sustainability, 2016, 8(5): 433.

[9] Cui Yanyan, Liu Lixin. Risk Evaluation of P2P Lending Platforms Based on Big Data [J]. Statistics & Information Forum, 2020, 35(04): 42-51.(In Chinese)

[10] Shi Rong, Zhang Te, Yang Guotao. Machine Learning Methods in Econometrics: Review and Prospects [J]. Statistics & Decision, 2024, 40(01): 52-56.(In Chinese)

[11] Tavana M., Abtahi A. R., Di Caprio D., et al. An Artificial Neural Network and Bayesian Network Model for Liquidity Risk Assessment in Banking [J]. Neurocomputing, 2018, 275: 2525-2554.

[12] Triepels R., Daniels H., Berndsen R. Monitoring Liquidity Management of Banks with Recurrent Neural Networks [J]. Computational Economics, 2021, 57(1): 89-112.

[13] Yang Junqi, Ren Rui, Kan Lina, et al. Research on Risk Assessment of Commercial Banks Based on BP Neural Network Model [J]. Friends of Accounting, 2021, (05): 113-119.(In Chinese)

[14] Tan Benyan, Gan Ziqi. Measurement and Early Warning of Commercial Bank Liquidity Risk Based on BO-LightGBM [J]. Wuhan Finance, 2022, (11): 21-32.(In Chinese)

[15] Sumi K V. Neural network-based liquidity risk prediction in Indian private banks[J]. Intelligent Systems with Applications, 2024, 21: 200322.