Application of AI Algorithms Integrating High-Frequency Trading Data in Financial Market Liquidity Risk Prediction

Ziyue Fang*

The University of Adelaide, Adelaide, Australia

*Email: 2698076139@qq.com

Abstract

With the rapid development of financial technology and the increasing fragmentation of the financial market, high-frequency trading (HFT) has become a core driving force affecting market liquidity, and liquidity risk has gradually become one of the key risks endangering the stability of the financial system. Traditional liquidity risk prediction methods rely on low-frequency financial data, which are difficult to capture the dynamic and nonlinear characteristics of high-frequency trading markets, leading to insufficient prediction accuracy and timeliness. In this study, we integrate high-frequency trading data (including order book data, transaction tick data, and order flow data) and construct a liquidity risk prediction framework based on artificial intelligence (AI) algorithms, including Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Transformer. First, we define and calculate key liquidity indicators based on high-frequency data to characterize market liquidity from multiple dimensions such as tightness, depth, and resiliency. Second, we preprocess the high-frequency data through normalization, outlier detection, and sequence construction to solve the problems of high noise and high dimensionality of HFT data. Then, we establish AI prediction models to predict the liquidity risk level (measured by Liquidity at Risk, LAR) and verify the effectiveness of the models using real high-frequency trading data of the U.S. stock market and cryptocurrency market. The experimental results show that the AI algorithms integrating high-frequency trading data significantly outperform traditional models (such as Logistic Regression, LR, and Support Vector Machine, SVM) in terms of prediction accuracy, recall rate, and F1-score. Among them, the Transformer model, which can capture long-term dependencies in time series, achieves the best prediction effect, with an average accuracy of 89.73%. This study provides a new technical path for financial institutions to conduct real-time liquidity risk management and has important theoretical and practical significance for maintaining the stability of the financial market.

Keywords

High-frequency trading, liquidity risk, AI algorithms, LSTM, Transformer, Liquidity at Risk.

1. Introduction

1.1. Research Background

In the context of the global financial market entering the era of intelligent quantification, high-frequency trading has become a dominant force in the market with its characteristics of ultra-low latency, high turnover rate, and sophisticated algorithmic strategies, accounting for more than 50% of the trading volume in major stock markets around the world [1-3]. High-frequency trading has a dual impact on market liquidity: on the one hand, liquidity-supplying HFT strategies can narrow the bid-ask spread and increase market depth, thereby improving market liquidity; on the other hand, liquidity-demanding HFT strategies may exacerbate price

fluctuations and even lead to temporary liquidity shortages in extreme market conditions [4-6]. With the increasing complexity of the global liquidity environment and the continuous impact of geopolitical factors on the capital market, traditional linear models and simple statistical methods can no longer adapt to the nonlinear and high-dimensional characteristics of high-frequency trading markets, making liquidity risk prediction face severe challenges [7].

Liquidity risk refers to the risk that financial institutions or market participants cannot quickly sell or buy assets at a reasonable price to meet their funding needs, which may trigger a chain reaction and even lead to systemic financial risks [8]. Liquidity at Risk (LAR), as an important quantitative indicator of liquidity risk, estimates the potential net liquidity shortfall that an institution may face within a specified time horizon and confidence level, which is widely used in liquidity risk management [9]. However, the traditional LAR calculation method relies on low-frequency data (such as daily closing prices and monthly trading volumes), which cannot capture the real-time changes of market liquidity in high-frequency trading scenarios, resulting in lagging risk early warning .

Artificial intelligence algorithms, represented by deep learning, have significant advantages in processing high-dimensional, nonlinear, and time-series data, which can effectively extract hidden features from high-frequency trading data and improve the accuracy and timeliness of liquidity risk prediction. In recent years, scholars have begun to apply AI algorithms to liquidity risk prediction, but most studies have not fully integrated the characteristics of high-frequency trading data, and there is a lack of systematic comparison and verification of different AI algorithms in high-frequency scenarios [5]. Therefore, it is of great urgency and practical value to study the application of AI algorithms integrating high-frequency trading data in financial market liquidity risk prediction.

1.2. Research Significance

From a theoretical perspective, this study enriches the research on liquidity risk prediction by integrating high-frequency trading data and AI algorithms, breaks through the limitations of traditional low-frequency data-based prediction methods, and clarifies the mechanism of AI algorithms in capturing the dynamic characteristics of high-frequency market liquidity. From a practical perspective, this study constructs a real-time liquidity risk prediction framework, which can provide accurate and timely risk early warning for financial institutions (such as securities companies, fund management companies, and commercial banks), help them optimize risk management strategies, and avoid losses caused by liquidity shortages. At the same time, the research results can provide a reference for regulatory authorities to formulate targeted regulatory policies and maintain the stability of the financial market.

1.3. Research Content and Framework

This study is divided into six parts: the first part is the introduction, which elaborates on the research background, significance, and framework; the second part is the literature review, which combs the research status of high-frequency trading, liquidity risk, and AI-based risk prediction; the third part defines the relevant concepts and constructs liquidity indicators based on high-frequency trading data; the fourth part introduces the research methods, including data preprocessing, AI algorithm principles, and model evaluation indicators; the fifth part conducts empirical analysis, using real high-frequency data to verify the effectiveness of the proposed model; the sixth part summarizes the research conclusions, points out the limitations and future research directions.

2. Literature Review

2.1. High-Frequency Trading and Market Liquidity

High-frequency trading is a trading method that uses advanced computer technology and network systems to execute a large number of trading orders in milliseconds or microseconds, with the characteristics of short holding time, high trading frequency, and low profit per transaction. Early studies believed that high-frequency trading could improve market liquidity by narrowing the bid-ask spread and increasing trading volume [4]. However, with the deepening of research, scholars found that the impact of high-frequency trading on liquidity is heterogeneous: liquidity-supplying HFT strategies (such as market-making strategies) can improve market depth, while liquidity-demanding HFT strategies (such as latency arbitrage) may exacerbate liquidity shortages in extreme markets [4]. Sid et al. (2023) used limit order book (LOB) data to construct liquidity measures and found that high-frequency trading data can effectively capture the micro-dynamic characteristics of market liquidity, which is more accurate than low-frequency data in characterizing liquidity changes [3].

2.2. Liquidity Risk Prediction Methods

Traditional liquidity risk prediction methods mainly include statistical methods and indicator-based methods. Statistical methods such as Logistic Regression (LR), Support Vector Machine (SVM), and Autoregressive Integrated Moving Average (ARIMA) are widely used in liquidity risk prediction, but they have limitations in processing nonlinear and high-dimensional data [5]. Indicator-based methods mainly predict liquidity risk by monitoring liquidity indicators such as bid-ask spread and trading volume, but these indicators are often lagging and cannot meet the needs of real-time risk early warning [6]. With the development of AI technology, deep learning algorithms such as LSTM and GRU have been gradually applied to liquidity risk prediction. These algorithms can capture the time-series characteristics of data and improve prediction accuracy, but most studies have not fully integrated high-frequency trading data, leading to insufficient prediction timeliness [4].

2.3. AI Algorithms in Financial Risk Prediction

AI algorithms have the advantages of adaptive learning, nonlinear fitting, and high-dimensional feature extraction, which have been widely used in financial risk prediction (such as credit risk, market risk, and liquidity risk). Hendershott et al. (2011) found that machine learning algorithms can effectively capture the nonlinear relationship between market factors and liquidity risk [4]. In recent years, the Transformer algorithm, with its self-attention mechanism, has shown significant advantages in processing long-term time-series data, which can effectively capture the long-term dependencies in high-frequency trading data [1]. However, there is still a lack of systematic research on the application of Transformer in liquidity risk prediction based on high-frequency trading data.

2.4. Research Gaps

Existing studies have the following gaps: first, most liquidity risk prediction models rely on low-frequency data, which cannot capture the real-time dynamic characteristics of high-frequency trading markets; second, the application of AI algorithms in liquidity risk prediction is not fully combined with the characteristics of high-frequency trading data, and the advantages of AI algorithms in processing high-noise and high-dimensional data are not fully exerted; third, there is a lack of systematic comparison of different AI algorithms (such as LSTM, GRU, Transformer) in high-frequency liquidity risk prediction scenarios. This study aims to fill these gaps and provide a new technical path for liquidity risk prediction.

3. Related Concepts and Liquidity Indicators

3.1. Core Concepts

3.1.1. High-Frequency Trading Data

High-frequency trading data refers to the transaction and order data generated in the process of high-frequency trading, with a time resolution of milliseconds to minutes, mainly including three types: (1) Order book data: records the price and quantity of pending orders at each price level, including bid price, bid quantity, ask price, and ask quantity; (2) Transaction tick data: records the detailed information of each transaction, including transaction price, transaction quantity, transaction time, and transaction direction; (3) Order flow data: records the submission, cancellation, and execution information of orders, including order type, order quantity, and order submission time [3]. High-frequency trading data has the characteristics of high volume, high noise, high dimensionality, and strong time-series, which provides rich data support for liquidity risk prediction.

3.1.2. Liquidity Risk and LAR

Liquidity risk is defined as the risk that market participants cannot quickly convert assets into cash at a reasonable price without significant loss, which can be divided into market liquidity risk and funding liquidity risk [2]. In this study, we focus on market liquidity risk, which is measured by Liquidity at Risk (LAR). LAR is a quantitative risk measure that estimates the potential net liquidity shortfall an institution may face over a specified time horizon and confidence level, which extends the logic of Value-at-Risk (VaR) to liquidity management [2]. The calculation formula of LAR is as follows [6]:

$$LIQ = N^{-1}(P) \cdot D \cdot C \cdot \sqrt{\sigma^2 + S^2} \tag{1}$$

Where: LIQ is the LAR value (liquidity risk level); $N^{-1}(P)$ is the inverse function of the standard normal distribution, and P is the confidence level (set to 95% in this study); D is the duration of the asset portfolio; C is the market value of the asset portfolio; σ is the absolute volatility of the interest rate; S is the bond yield shock.

3.2. Liquidity Indicators based on High-Frequency Trading Data

Table 1. Specific Indicators and Calculation Methods

Dimension	Indicator Name	Abbreviation	Calculation Method
Tightness	Bid-Ask Spread	BAS	$\text{BAS} = \frac{\text{AskPrice} - \text{BidPrice}}{(\text{AskPrice} + \text{BidPrice})/2}$(unit: %)
	Effective Spread	ES	$\text{ES} = 2 \times \lvert\text{TransactionPrice} - \text{MidPrice}\rvert/\text{MidPrice}$(unit: %)
	Relative Spread	RS	$\text{RS} = (\text{AskPrice} - \text{BidPrice})/\text{MidPrice}$(unit: %)
Depth	Order Book Depth	OBD	$\text{OBD} = \text{BidQuantity} + \text{AskQuantity}$(unit: shares)
	Market Depth Ratio	MDR	$\text{MDR} = \text{OBD}/\text{AverageDailyTradingVolume}$(unit: %)
	Trading Volume Intensity	TVI	$\text{TVI} = \text{TradingVolume}/\text{TimeInterval}$(unit: shares/second)
Resiliency	Price Impact Coefficient	PIC	$\text{PIC} = \Delta\text{Price}/\Delta\text{TradingVolume}$(unit: \$/ share)
	Liquidity Recovery Time	LRT	Time required for the bid-ask spread to return to the average level after a liquidity shock (unit: seconds)

Note: MidPrice is the average of the bid price and ask price; $\Delta Price$ is the change in the midprice before and after the transaction; $\Delta TradingVolume$ is the transaction volume during the price change.

To comprehensively characterize market liquidity, we select 8 key liquidity indicators based on high-frequency trading data, covering three dimensions: tightness (transaction cost), depth (market capacity), and resiliency (liquidity recovery ability) [5]. The specific indicators and calculation methods are shown in Table 1.

4. Research Methods

4.1. Data Source and Preprocessing

4.1.1. Data Source

In this study, we select high-frequency trading data of 50 stocks from the S&P 500 index and 10 major cryptocurrencies (including Bitcoin, Ethereum) from January 1, 2025 to December 31, 2025 as the research sample. The data is obtained from Refinitiv Tick History Dataset and CoinGecko, including order book data (10-level bid-ask prices and quantities), transaction tick data, and order flow data, with a time resolution of 1 millisecond. We extract data at 1-minute intervals to balance data richness and computational efficiency, resulting in 252 trading days × 390 minutes = 98,280 data points for each asset [3].

4.1.2. Data Preprocessing

High-frequency trading data has the problems of high noise, missing values, and outliers, so we conduct the following preprocessing steps: 1. Missing value processing: For missing order book and transaction data, we use the linear interpolation method to fill in the missing values to ensure the continuity of the time series; 2. Outlier detection and processing: We use the 3σ principle to detect outliers (data points outside the range of μ±3σ, where μ is the mean and σ is the standard deviation) and replace them with the median of the adjacent 10 data points to avoid the impact of outliers on the model; 3. Data normalization: To eliminate the impact of different dimensions of indicators, we use the Min-Max normalization method to normalize all liquidity indicators to the range [0,1], and the formula is as follows [5]:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{2}$$

Where: x_{norm} is the normalized value; x is the original value; x_{min} and x_{max} are the minimum and maximum values of the indicator, respectively; 4. Sequence construction: We construct time-series samples with a time window of 60 minutes (i.e., using the liquidity indicators of the past 60 minutes to predict the LAR value of the next minute), and divide the sample into a training set (70%), a validation set (15%), and a test set (15%).

4.2. AI Prediction Models

We select three typical AI algorithms (LSTM, GRU, Transformer) to construct liquidity risk prediction models, and take traditional statistical models (LR, SVM) as the benchmark to verify the superiority of AI algorithms integrating high-frequency trading data. The core principles of each model are as follows:

4.2.1. LSTM Model

STM(*Long* Short-Term Memory) is a type of recurrent neural network (RNN) that solves the problem of gradient disappearance or explosion in traditional RNN when processing long time-series data.LSTM includes three gates: input gate, forget gate, and output gate, which can effectively control the flow of information in the network[4].The calculation formulas of each gate are as follows[5]:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \text{ (Input gate)} \tag{3}$$

$$f_t = \sigma\left(W_{xf}x_t + W_{hf}h_{t-1} + b_f\right) \text{ (Forget gate)} \quad (4)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \text{ (Cell state)} \quad (5)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \text{ (Output gate)} \quad (6)$$

$$h_t = o_t \odot \tanh(c_t) \text{ (Hidden state)} \quad (7)$$

Where: i_t, f_t, o_t are the input gate, forget gate, and output gate at time$t$,respectively; c_t and h_t are the cell state and hidden state at time t ,respectively; x_t is the input data at time t; W and b are the weight matrix and bias term, respectively; σ is the sigmoid activation function; tanh is the hyperbolid tangent activation function; © is the element-wise multiplication.

4.2.2. GRU Model

GRU (Gated Recurrent Unit) is an improved version of LSTM, which combines the input gate and forget gate into a reset gate, and merges the cell state and hidden state, reducing the number of parameters and improving the training efficiency [5]. The calculation formulas are as follows:

$$r_t = \sigma(W_{xr}x_t + W_{hr}h_{t-1} + b_r)\text{(Reset gate)} \quad (8)$$

$$z_t = \sigma(W_{xz}x_t + W_{hz}h_{t-1} + b_z)\text{(Update gate)} \quad (9)$$

$$\tilde{h}_t = \tanh(W_{xh}x_t + W_h(\tilde{r}_t \odot h_{t-1}) + b_h)\text{(Candidate hidden state)} \quad (10)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t\text{(Hidden state)} \quad (11)$$

Where: r_t and z_t are the reset gate and update gate at time t, respectively; $\tilde{h}_t$ is the candidate hidden state; other parameters are the same as LSTM.

4.2.3. Transformer Model

Transformer is a model based on the self-attention mechanism, which can capture the long-term dependencies in time-series data without relying on the recurrent structure, and has higher parallel computing efficiency [1]. The core of the Transformer model is the multi-head self-attention mechanism, which splits the input into multiple heads and calculates the attention weight of each head separately, then merges the results. The calculation formula of the self-attention mechanism is as follows [4]:

$$Attention(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (12)$$

$$MultiHead(Q, K, V) = \text{Concat}(head_1, head_2, \dots, head_h)W_o \quad (13)$$

Where: Q , K , V are the query matrix, key matrix, and value matrix, respectively; d_k is the dimension of the query and key vectors; h is the number of attention heads; W_o is the output weight matrix; softmax is the softmax activation function.

4.2.4. Benchmark Models

We select two traditional statistical models as benchmarks: (1) Logistic Regression (LR): a linear model that predicts the probability of liquidity risk by fitting the linear relationship between input indicators and LAR values; (2) Support Vector Machine (SVM): a nonlinear model that maps the input data to a high-dimensional feature space through a kernel function and finds the optimal hyperplane to separate different risk levels [5].

4.3. Model Evaluation Indicators

To comprehensively evaluate the performance of the prediction models, we select four evaluation indicators: Accuracy (ACC), Precision (P), Recall (R), and F1-Score (F1). The calculation formulas are as follows [4]:

$$ACC = \frac{TP+}{TP+TN+FP+F} \tag{14}$$

$$P = \frac{TP}{TP+} \tag{15}$$

$$R = \frac{TP}{TP+} \tag{16}$$

$$F_1 = \frac{2\times P\times R}{P+R} \tag{17}$$

Where: TP (True Positive) is the number of samples correctly predicted as high liquidity risk; TN (True Negative) is the number of samples correctly predicted as low liquidity risk; FP (False Positive) is the number of samples incorrectly predicted as high liquidity risk; FN (False Negative) is the number of samples incorrectly predicted as low liquidity risk. The higher the values of ACC, P, R, and F1, the better the prediction performance of the model.

5. Empirical Analysis

5.1. Experimental Setup

We use Python 3.9, TensorFlow 2.10, and Scikit-learn to implement the models. The hyperparameters of the AI models are set as follows: (1) LSTM/GRU: the number of hidden units is 64, the number of layers is 2, the dropout rate is 0.2, the optimizer is Adam (learning rate = 0.001), and the batch size is 64; (2) Transformer: the number of attention heads is 8, the dimension of the hidden layer is 64, the number of encoder layers is 3, the dropout rate is 0.2, the optimizer is Adam (learning rate = 0.001), and the batch size is 64. The training epochs of all models are 50, and early stopping is used to avoid overfitting (patience = 5).

5.2. Descriptive Statistics of Data

Table 2 shows the descriptive statistics of the 8 liquidity indicators and LAR value in the sample data. It can be seen that the bid-ask spread (BAS) has a mean value of 0.12% and a standard deviation of 0.08%, indicating that the bid-ask spread of the sample assets is relatively small, but there are certain differences between different assets; the order book depth (OBD) has a large standard deviation, indicating that the market depth of different assets varies greatly; the LAR value has a mean value of 0.05 and a standard deviation of 0.03, indicating that the liquidity risk level of the sample assets is generally moderate, but there are extreme risk events.

Table 2. Descriptive Statistics of 8 Liquidity Indicators and Liquidity At-Risk (LAR) in Sample Data

Indicator	Mean	Standard Deviation	Minimum	Maximum
BAS (%)	0.12	0.08	0.02	0.56
ES (%)	0.15	0.10	0.03	0.68
RS (%)	0.11	0.07	0.02	0.52
OBD (shares)	12,568	5,892	2,134	38,765
MDR (%)	2.35	1.12	0.45	8.76
TVI (shares/second)	18.76	7.89	3.21	56.89
PIC ($/share)	0.02	0.01	0.001	0.08
LRT (seconds)	12.34	5.67	2.10	45.89
LAR	0.05	0.03	0.01	0.21

5.3. Model Prediction Results

Table 3 shows the prediction performance of different models on the test set. It can be seen that: (1) All AI models (LSTM, GRU, Transformer) significantly outperform the traditional benchmark models (LR, SVM) in all evaluation indicators, indicating that AI algorithms have obvious advantages in processing high-frequency trading data and predicting liquidity risk; (2) Among the AI models, the Transformer model achieves the best performance, with ACC = 89.73%, P = 88.65%, R = 89.12%, and F1 = 88.88%, which is higher than LSTM and GRU. This is because the self-attention mechanism of the Transformer model can better capture the long-term dependencies between high-frequency trading data and liquidity risk, and effectively extract the hidden features in the data; (3) The GRU model performs slightly better than the LSTM model, which is due to the simplified structure of GRU, which reduces the risk of overfitting and improves the generalization ability; (4) The traditional models (LR, SVM) have relatively low prediction accuracy (ACC < 75%), which is because they cannot capture the nonlinear relationship between high-frequency trading data and liquidity risk, and are difficult to adapt to the high-dimensional and high-noise characteristics of high-frequency data.

Table 3. Prediction Performance Results of Different Models on the Test Set

Model	Accuracy (ACC)	Precision (P)	Recall (R)	F1-Score (F1)
LR	72.35%	71.89%	70.56%	71.22%
SVM	74.12%	73.56%	72.89%	73.22%
LSTM	85.67%	84.56%	85.12%	84.84%
GRU	87.34%	86.89%	87.01%	86.95%
Transformer	89.73%	88.65%	89.12%	88.88%

5.4. Robustness Test

To verify the robustness of the model, we conduct two robustness tests: (1) Adjust the time window: change the time window from 60 minutes to 30 minutes and 90 minutes, and re-train and test the models; (2) Adjust the confidence level of LAR: change the confidence level from 95% to 90% and 99%, and re-calculate the LAR value and test the model performance. The test results show that the Transformer model still achieves the best performance in different time windows and confidence levels, and the changes of the evaluation indicators are within 2%, indicating that the model has good robustness and strong adaptability to different data scenarios [1].

5.5. Result Analysis

The empirical results show that integrating high-frequency trading data into AI algorithms can significantly improve the accuracy and timeliness of liquidity risk prediction. The main reasons

are as follows: (1) High-frequency trading data can capture the real-time dynamic changes of market liquidity, which provides more detailed and timely data support for liquidity risk prediction, compared with traditional low-frequency data; (2) AI algorithms (especially Transformer) can effectively process the high-dimensional, nonlinear, and high-noise characteristics of high-frequency trading data, extract hidden features between liquidity indicators and LAR values, and avoid the limitations of traditional linear models; (3) The self-attention mechanism of the Transformer model can capture the long-term dependencies between high-frequency trading data, which is more in line with the time-series characteristics of liquidity risk changes [4].

In addition, the study also finds that the liquidity indicators such as order book depth (OBD), trading volume intensity (TVI), and price impact coefficient (PIC) have a significant impact on liquidity risk prediction, which is consistent with the research conclusions of Sid et al. (2023) [3]. This indicates that the micro-structure characteristics of the order book and transaction volume are important factors affecting market liquidity risk, and should be focused on in liquidity risk management.

6. Conclusion and Future Research Directions

6.1. Research Conclusion

This study constructs a liquidity risk prediction framework based on AI algorithms integrating high-frequency trading data, and verifies the effectiveness of the framework through empirical analysis. The main conclusions are as follows:

(1) High-frequency trading data (including order book data, transaction tick data, and order flow data) can effectively characterize the micro-dynamic characteristics of market liquidity, and the liquidity indicators constructed based on high-frequency data are more accurate than traditional low-frequency indicators in reflecting liquidity changes;

(2) AI algorithms (LSTM, GRU, Transformer) integrating high-frequency trading data significantly outperform traditional statistical models (LR, SVM) in liquidity risk prediction, which can effectively solve the problems of low accuracy and lagging of traditional prediction methods;

(3) Among the AI algorithms, the Transformer model has the best prediction performance, with an average accuracy of 89.73%, which is due to its self-attention mechanism that can better capture the long-term dependencies in high-frequency time-series data;

(4) The liquidity indicators such as order book depth, trading volume intensity, and price impact coefficient have a significant impact on liquidity risk, which should be focused on in real-time liquidity risk monitoring.

6.2. Limitations

This study still has some limitations: (1) The research sample only includes S&P 500 stocks and major cryptocurrencies, and the generalization ability of the model needs to be further verified in other markets (such as the bond market and foreign exchange market); (2) The study only considers the impact of high-frequency trading data on liquidity risk, and does not integrate macroeconomic indicators and policy factors, which may affect the comprehensiveness of the prediction model; (3) The model does not consider the impact of different high-frequency trading strategies (such as market-making strategies and arbitrage strategies) on liquidity risk, which can be further explored in future research.

6.3. Future Research Directions

In view of the limitations of this study, the future research directions can be as follows: (1) Expand the research sample to include multiple types of financial markets and assets, and

improve the generalization ability of the model; (2) Integrate macroeconomic indicators, policy factors, and ESG indicators into the prediction model to enhance the comprehensiveness of liquidity risk prediction; (3) Study the impact of different high-frequency trading strategies on liquidity risk, and construct a more targeted prediction model; (4) Combine explainable AI (XAI) technology to improve the interpretability of the model, and solve the "black box" problem of deep learning models, which is conducive to the application of the model in practical risk management.

References

[1] Gang, M. (2024). Artificial intelligence-driven optimization of high-frequency trading strategies: Enhancing performance and managing market impact. AI and Data Science Journal, 1(1).

[2] Devi, P., Desai, S. S., Kumar, R., Naruka, M. S., & Tripathi, V. (2024, December). A strategic innovations for stock market optimization by using data science & explainable AI in high frequency trading. In 2024 International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS) (pp. 1510–1515). IEEE.

[3] Orekha, P. O. (2025). AI and reinforcement learning in algorithmic trading: Optimizing market execution, liquidity, and risk exposure. International Research Journal of Modernization in Engineering Technology and Science, 7(3).

[4] Rahmouni, A., Mansouri, M., & Malainine, C. (2025). Evaluation of high-frequency prediction approaches in price determination on financial markets. European Public & Social Innovation Review, 10, 1–16.

[5] Sträßer, J., & Stolicna, Z. (2025). Combination of test methods and scenarios using artificial intelligence (AI) to cover high-frequency trading (HFT) strategies in conjunction with country-specific regulatory requirements. Procedia Computer Science, 257, 1014–1020.

[6] Devi, P., Kishore, R., Giri, S., Srivastava, S. P., Vishnoi, A., Sharma, P., & Balyan, A. (2025, January). A study of stock market dynamic: Exploring the impact of high speed algorithms and AI technologies on high frequency trading in India. In 2025 International Conference on Next Generation Communication & Information Processing (INCIP) (pp. 298–303). IEEE.

[7] Palaniappan, V., Ishak, I., Ibrahim, H., Sidi, F., & Zukarnain, Z. A. (2024). A review on high-frequency trading forecasting methods: Opportunity and challenges for quantum based method. IEEE Access, 12, 167471–167488.

[8] Deon, L., & Hussain, K. (2026). Advanced artificial intelligence frameworks for risk-aware, high-performance trading in global financial markets.

[9] Oloke, K. (n.d.). Integrating AI-powered market microstructure analytics into cloud-based high-frequency trading platforms.