Proceedings of the 7th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/2027.38689

# *Intelligent Detection of Securities Market Manipulation Using Multimodal Financial Data*

**Yanying Li**

*The Chinese University of Hong Kong, Hong Kong, China*
*liyanying02605@outlook.com*

***Abstract.*** Market manipulation through securities involves increasingly sophisticated mechanisms such as irregular trading, selective information disclosures, and joint activities in different securities, posing threats to surveillance relying solely on a particular information source due to both false positives and late identification of manipulations. We propose a multimodal framework for the detection of market manipulation on China's A-share market by fusing one-minute market microstructure time series, corporate announcements, financial news, and evolving relationships among securities. Enforced cases by China Securities Regulatory Commission from 2018–2025 were combined with unenforced stock-period observations and temporal, textual, and relational features are extracted using Temporal Convolutional Network, MacBERT encoder, and Graph Attention Network, respectively, followed by cross-modal attention to generate a risk score. Our framework achieved macro-F1 of 0.872 ± 0.009, AUROC of 0.946 ± 0.006, and AUPRC of 0.821 ± 0.012 on the unseen data, outperforming the TCN-only model in market only by a wide margin of 0.036 ± 0.007 and 0.028 ± 0.006 for removing text and graph information, respectively. In the 2025 market surveillance simulation, reviewing the top 10% of ranked observations led to detecting 0.643 ± 0.031 of manipulations with 0.074 ± 0.009 false-positive rate.

***Keywords:*** Securities Market Manipulation, Multimodal Financial Data, Market Surveillance, Cross-Modal Learning, Explainable Detection

## 1. Introduction

Market manipulation continues to be hard to monitor due to abnormal prices formation being spread across orders, trades, information, and manipulation among related securities. There is empirical evidence showing that learning models can detect manipulation better when machine-learning models are trained with conventional market variables complemented by sentiments and information indicators [1]. A literature survey for trade-based manipulation detection reveals that heterogeneous anomalies, changing manipulation strategies, and low number of labeled cases continue to form the core of methodological problems [2, 3].

The rising availability of high-frequency data encourages the development of adaptive approaches. Pattern recognition models that use large sets of order book data can differentiate manipulative behavior from normal market behavior without having to rely on rigid thresholds set

© 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: 10.54254/2755-2721/2027.38689

by supervisors [4]. Deep anomaly detection approaches reveal the ability of learning time representations from limit order book data to detect quote stuffing, layering, and other anomalies in trades out-of-sample [5]. Ensemble deep learning methods have further improved stock price manipulation detection through preserving the information contained in many submodels that represent time instead of having one handcrafted set of indicators [6]. Jump-based deep learning of high-frequency data also reveals manipulation signals contained in abnormal price dynamics [7]. Trading behavior, however, is not the only factor of the surveillance context. Financial news has been found capable of significantly improving stock-state representation through encoding textual information along with market signals [8]. Multimodal graphs have revealed that there are complementary dependencies between textual, temporal, and cross-security dependencies in financial inference [9]. Market, textual, and graph features have also been able to enhance extreme stock-risk state detection [10]. On this basis, this study represents manipulation as a multimodal anomaly recognition task. The cross-modal attention mechanism fuses high-frequency microstructure sequences, public information, and dynamic cross-security relations, resulting in risk scores and modality-level evidence for regulation.

## 2. Literature review

### 2.1. Trade-based market manipulation and observable trading anomalies

Manipulation based on trade activity may distort the usual interrelationship between prices, volumes, liquidity, and order flow. Signals are characterized by abnormal returns, concentration of trades, aggression of order imbalance, bursts of cancellations, intraday reversals, short-term volatility, and liquidity withdrawals. Such features are informative since manipulation usually involves multiple transactions or order placements. The same signals can be observed after information shock or in periods of market-wide stress [1-4]. Thus, reliable manipulation detection requires not only consideration of thresholds but also sequential context and comparison with the surrounding information environment.

### 2.2. Machine-learning models for securities market surveillance

Machine-learning surveillance substitutes rule-based systems by nonlinear mapping from the state of the market to the level of manipulation risks. In this case, tree-based models are still useful for tabular indicators. Recurrent neural networks, as well as convolutional and transformer-based architectures, are capable of preserving the order of emergence of abnormal events. The time dimension is especially important in cases of short manipulation windows when order imbalance, volume acceleration, and reversals take place within minutes. Label scarcity becomes critical since the outcomes of enforcement procedure is a part of manipulation instances. Therefore, robust surveillance implies chronological validation, imbalance awareness, and explicit management of false positives [5-7].

### 2.3. Multimodal evidence of manipulation

The suspicious trading activity is more convincing if the abnormal behavior of the market is unrelated to contemporaneous public information and anomalies spread across economically or statistically related securities. The text contains the information on whether price changes are related

DOI: 10.54254/2755-2721/2027.38689

to announcement or news, while the graph structure reflects the synchronized price movements that cannot be captured in stock-level models [8-10].

## 3. Methodology

### 3.1. Dataset construction and manipulation event labeling

The sample contains Chinese A-shares securities between 2018 and 2025. Confirmed instances of manipulations are sourced from the China Securities Regulatory Commission administrative decisions and match the security codes, manipulations time ranges, and the enforcement dates. Every positive instance of a stock is matched against four negative instances that share the same industry as well as calendar year based on the stock's market cap and realized volatility in the pre-window period. Window period is defined as 20 days before and 10 days after the event center date. In order to prevent any data leakage, enforcement documents and media coverage after enforcement are not used as inputs for the models.

### 3.2. High-frequency market microstructure features

The one-minute transaction and quote data of CSMAR are mapped into the intraday sequence of the same stock-month. The core features are log return, turnover, quoted spread, volume imbalance, book imbalance, realized volatility, cancellation ratio, depth pressure, volume acceleration, and five-minute reversal. Every feature is standardized within each stock-month in order to control for persistency of liquidity. A causal Temporal Convolutional Network with four dilated residual blocks, kernel size three, and dilation levels 1, 2, 4, and 8, provides the network with an adequate width of the receptive field without resorting to the information about the future. The residual channel dimension is set to 128 with dropout equal to 0.20.

### 3.3. Financial text representation

Announcements and financial news from listed firms are aligned according to their stock code and time stamp of publication. Documents that are published after the closing date of the observation period are not considered in the study. The corpus is then encoded through MacBERT model. Attention pooling is employed to create a vector of information for each day. Four additional features are created based on the encoding process of the corpus; namely, sentiment intensity, uncertainty, event novelty, and semantic consistency between announcement and financial news. The gap in information is estimated using the comparison between standardized market activity intensity and maximum contemporary material information intensity.

### 3.4. Cross-security relationship modeling

For each week, the graph is dynamically generated, taking securities as nodes. The edge will be formed between the two stocks if the stocks lie in the same CSRC industry or the residual return correlation for the last 60 days is above the 95th percentile of the market distribution. The edge weight consists of the affinity of the industry and correlation value. The Graph Attention Network with two layers is used to aggregate the neighboring abnormal returns, liquidity, and volume states in 128 dimensions using 8 attention heads. With the attention values, it is possible to focus on unusual synchronization neighbors instead of all industry connections.

DOI: 10.54254/2755-2721/2027.38689

### 3.5. Cross-modal manipulation detection model

The temporal, textual, and graph embeddings are projected into a common 256-dimensional space. Cross-modal attention learns event-specific evidence weights, and a gated residual connection preserves modality-specific signals when one source is weak. For stock-period i, the fused representation and manipulation probability are defined as shown in Equation (1):

$$z_i = \sum_{m\in\{M,T,G\}} \frac{exp(q_i^T W_m h_i^{(m)}/\sqrt{d})}{\sum_{j\in\{M,T,G\}} exp(q_i^T W_j h_i^{(j)}/\sqrt{d})} W_m h_i^{(m)} + \gamma(h_i^{(M)} \odot g_i) \quad (1)$$

Here, $M, T, G$ denote market, text, and graph modalities. The learned attention coefficients are retained as modality-level evidence. SHAP is applied to the underlying microstructure features, while linked news, announcements, and high-attention neighboring securities are returned with each high-risk alert. The event-level multimodal evidence fusion process is illustrated in Figure 1.

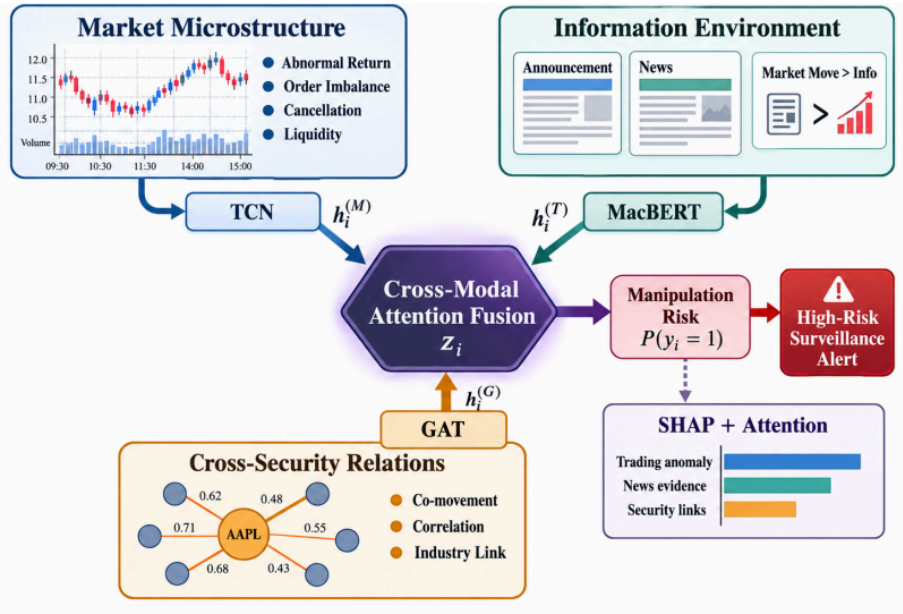



Figure 1. Event-level multimodal evidence fusion for securities market manipulation detection

## 4. Experimental design and validation

### 4.1. Strategy for training and baseline comparison

Observations from 2018-2022 are used for training, those from 2023 for validation, and observations from 2024-2025 are withheld as holdout. Training is done with AdamW with learning rate (2x10^-5), batch size 64, and early stopping on validation AUPRC. Class-weighted focal loss is used to address imbalanced classes. Logistic regression, XGBoost, a market-only TCN, and a fusion market-and-text model are the comparators. Results are provided as precision, recall, macro-F1, AUROC, AUPRC, and false positive rate for five random seeds.

DOI: 10.54254/2755-2721/2027.38689

### 4.2. Ablation and robustness studies

Ablation studies independently omit text, graph relation, and order-book features. Robustness studies analyze the performance over terciles of realized volatility and market capitalization. A 30 day forward shift study shifts each test window 30 days forward without re-training to analyze temporal degradation. Statistical differences are analyzed via paired bootstrap resampling, with 2,000 replications. In addition, calibration is analyzed with the Brier score to avoid improved ranking from hiding instability in probability estimates.

### 4.3. Regulatory surveillance simulation

The held-out observations in 2025 are ranked according to the predicted risk of manipulation. Simulated budgets of review for regulators are provided for the top 5%, 10%, and 20% of observations versus random selection of those same budget amounts. The interface provides for each alert the microstructure features which are most indicative, public text, graph neighbor influence, modalities used and SHAP values. Efficiency is measured as the number of confirmed manipulation events per reviewed observation, along with the false positive burden.

## 5. Results and discussion

### 5.1. Detection performance and incremental value of multimodal financial data

The complete model achieved macro-F1 of 0.872 ± 0.009***, AUROC of 0.946 ± 0.006***, and AUPRC of 0.821 ± 0.012***. Relative to the market-only TCN, the full system improved macro-F1 by 0.071 ± 0.011 and reduced the false-positive rate from 0.119 ± 0.012 to 0.074 ± 0.009 (see Table 1). Adding text to market data produced the largest single-stage precision gain, whereas graph information contributed most strongly to recall among coordinated-event observations.

Table 1. Detection performance of the proposed multimodal manipulation model and baselines

| Model | Precision | Recall | Macro-F1 | AUROC | AUPRC | FPR |
|---|---|---|---|---|---|---|
| Proposed multimodal model | 0.889 ± 0.011*** | 0.858 ± 0.014*** | 0.872 ± 0.009*** | 0.946 ± 0.006*** | 0.821 ± 0.012*** | 0.074 ± 0.009*** |
| Market + text | 0.854 ± 0.014** | 0.817 ± 0.017** | 0.834 ± 0.012** | 0.918 ± 0.008** | 0.777 ± 0.015** | 0.091 ± 0.010** |
| Market-only TCN | 0.826 ± 0.016* | 0.781 ± 0.020* | 0.801 ± 0.014* | 0.892 ± 0.010* | 0.724 ± 0.018* | 0.119 ± 0.012 |
| XGBoost | 0.781 ± 0.019 | 0.738 ± 0.022 | 0.758 ± 0.016 | 0.851 ± 0.012 | 0.669 ± 0.021 | 0.143 ± 0.014 |
| Logistic regression | 0.702 ± 0.021 | 0.661 ± 0.025 | 0.680 ± 0.019 | 0.793 ± 0.015 | 0.588 ± 0.024 | 0.181 ± 0.017 |

**Paired bootstrap comparison with logistic regression: $p < 0.05$, *$p < 0.01$, *$p < 0.001$.*

DOI: 10.54254/2755-2721/2027.38689

### 5.2. Patterns of manipulation risks across trading, information, and relational modalities

Manipulation window manipulations had standardized order-book imbalance of 1.72 ± 0.38 compared to 0.61 ± 0.27 in control groups. Information gap scores were 1.46 ± 0.33 and 0.52 ± 0.24 for confirmed and control events. Concentration in graph-attention was 0.684 ± 0.071 in manipulation risks and 0.397 ± 0.064 in control groups. Removal of text lowered macro-F1 by 0.036 ± 0.007**, while removing graphs lowered it by 0.028 ± 0.006**, thus proving unexplained price movements and coordination of relations provided additional independent information apart from microstructure information.

### 5.3. Efficiency of regulatory screening, model explainability, and control of false positives

The review of the top 10% of 2025 ranked events captured 0.643 ± 0.031*** of confirmed events compared to 0.108 ± 0.018 with random screening. With the budget of 20% of reviews, capture was 0.817 ± 0.024, while FPR was kept at 0.096 ± 0.011. For the 30-day forward shift experiment, AUROC was 0.921 ± 0.008 and AUPRC was 0.786 ± 0.015. Modality and SHAP explanations showed good stability across the seeds with mean top-five features overlap of 0.842 ± 0.029. False positive were mostly clustered around earnings announcement and sectorwide volatility shock events, which confirms necessity of human review of alert generation.

## 6. Conclusion

In this paper, we develop an explainable multimodal approach to intelligent detection of securities-market manipulation. By considering high frequency microstructure sequences, text and dynamic cross-security relation networks, our model distinguishes isolated volatile events from suspicious patterns having support from multiple evidence modalities. Cross-modal attention increases detectability and preserves modality traceability, and regulation simulation proves effectiveness of risk ranking to concentrate confirmed cases using limited regulatory review capacity. Our approach is designed for use in surveillance prioritization: final conclusions of enforcement actions require analysis of trading intentions, account relationships, transaction history and relevant regulations.

## References

[1] Liu, Q., Wang, C., Zhang, P., & Zheng, K. (2021). Detecting stock market manipulation via machine learning: Evidence from China Securities Regulatory Commission punishment cases. *International Review of Financial Analysis, 78*, 101887.

[2] Khodabandehlou, S., & Hashemi Golpayegani, S. A. (2022). Market manipulation detection: A systematic literature review. *Expert Systems with Applications, 210*, 118330.

[3] Zulkifley, M. A., Munir, A. F., Sukor, M. E. A., & Shafiai, M. H. M. (2023). A survey on stock market manipulation detectors using artificial intelligence. *Computers, Materials & Continua, 75*(2), 4395–4418.

[4] James, R., Leung, H., & Prokhorov, A. (2023). A machine learning attack on illegal trading. *Journal of Banking & Finance, 148*, 106735.

[5] Poutré, C., Chételat, D., & Morales, M. (2024). Deep unsupervised anomaly detection in high-frequency markets. *The Journal of Finance and Data Science, 10*, 100129.

[6] Liu, C., Li, S., & Shi, L. (2024). A stock price manipulation detecting model with ensemble learning. *Expert Systems with Applications, 248*, 123479.

[7] Huang, H., & Wu, Y. (2026). Deep learning-based high-frequency jump test for detecting stock market manipulation: Evidence from China's securities market. *Kybernetes, 55*(1), 658–674.

[8] Chen, X., Ma, X., Wang, H., Li, X., & Zhang, C. (2022). A hierarchical attention network for stock prediction based on attentive multi-view news learning. *Neurocomputing, 504*, 1–15.

DOI: 10.54254/2755-2721/2027.38689

[9] Liu, R., Liu, H., Huang, H., Song, B., & Wu, Q. (2024). Multimodal multiscale dynamic graph convolution networks for stock price prediction. *Pattern Recognition, 149*, 110211.
[10] Sheng, Y., Qu, Y., & Ma, D. (2024). Stock price crash prediction based on multimodal data machine learning models. *Finance Research Letters, 62*, 105195.