Article

Spatiotemporal Risk Representation Learning Using Transformers and Graph Structure

Xinyi Liang , Yinghao Zhao , Mingfan Chang , Ruizhe Zhou , Kewei Cao , Yihan Zheng *

Posted Date: 17 March 2026

doi: 10.20944/preprints202603.1285.v1

Keywords: spatiotemporal modeling; financial crisis warning; attention mechanism; risk identification

doi:10.20944/preprints202603.1285.v1

Article

Spatiotemporal Risk Representation Learning Using Transformers and Graph Structure

Xinyi Liang [1], Yinghao Zhao [2], Mingfan Chang [3], Ruizhe Zhou [4], Kewei Cao [1] and Yihan Zheng [1,*]

[1] Columbia University, New York, USA
[2] Pace University, New York, USA
[3] Brandeis University, Waltham, USA
[4] University of Chicago, Chicago, USA
* Correspondence: Gary.zheng22@gmail.com

Abstract

This study addresses the challenges of financial risk early warning by proposing a modeling approach based on spatiotemporal Transformers. The research first examines the multidimensional characteristics of financial risk, emphasizing its temporal dynamics and cross-regional interactions. It notes that many existing methods struggle to jointly capture temporal dependencies and inter-regional risk transmission patterns. To overcome these limitations, a unified spatiotemporal modeling framework is developed. The framework integrates temporal encoding, spatial adjacency information, and multi-head attention mechanisms to model long-range dependencies and regional spillover effects. In the model architecture, an embedding layer is employed to learn representations from multi-source financial indicators. A self-attention mechanism facilitates global feature interaction, while a graph convolution component further enhances the modeling of spatial relationships across markets. The final risk representation is generated through a feed-forward network with normalization layers, providing a structured basis for financial risk assessment and early warning analysis. Experimental evaluations include comparative studies and sensitivity analyses under varying missing data ratios, time window settings, and environmental conditions. The results indicate that the proposed method consistently outperforms several baseline models in terms of accuracy, precision, recall, and F1-score. Overall, the approach demonstrates strong robustness and practical applicability in complex financial settings, offering an effective tool for financial risk monitoring and decision support.

Keywords: spatiotemporal modeling; financial crisis warning; attention mechanism; risk identification

CCS Concepts: computing methodologies; machine learning; machine learning approaches

---

I. Introduction

In the context of globalization and deep financialization, the stability of financial systems has become a core concern for governments and societies worldwide. Financial crises are often accompanied by sharp fluctuations in asset prices, imbalances in credit markets, and cross-regional contagion effects. These crises exert profound impacts on economic growth and social welfare [1]. In today's highly interconnected economy, local market volatility can quickly escalate into systemic crises through capital flows, trade dependencies, and the spread of information. Therefore, developing advanced technologies to achieve early identification and warning of financial risks is not only an important research issue but also a strategic requirement for safeguarding national economic security and social stability [2].

Against this backdrop, the spatiotemporal characteristics of financial risk have received increasing attention from both academia and industry. Risks in financial markets do not exist in

doi:10.20944/preprints202603.1285.v1

isolation. Instead, they evolve and spread across space [3]. For example, adjustments in macroeconomic policy, international market linkages, regional capital flows, and structural imbalances across industries all contribute to the dynamic clustering and transmission of risk. These multi-dimensional interactions reveal that single-dimension or static methods are insufficient to capture the internal logic of crisis formation. Frameworks that model both temporal dependencies and spatial correlations align more closely with the essential patterns of financial risk evolution [4].

With the rise of artificial intelligence and deep learning, spatiotemporal modeling techniques have been introduced into financial risk research [5]. Among these, the Transformer architecture has drawn wide attention for its strengths in handling long-term dependencies and extracting global features. Traditional time series methods struggle with nonlinear and multi-scale fluctuations, while attention-based models can capture long-range dependencies and complex interactions. When spatial dimensions are incorporated into the framework, spillover effects across markets, heterogeneous transmissions across regions, and resonance across industries can all be represented within a unified spatiotemporal space. This provides new theoretical and methodological support for early warning of financial crises.

At the application level, early warning of financial crises is not only a matter of academic inquiry but also a critical foundation for public policy and regulatory practice. Efficient and accurate warning systems can provide governments with timely intervention tools and reduce the social costs of large economic swings. Such systems play an essential role in improving macroeconomic stability by helping authorities monitor systemic risks, assess vulnerabilities, and act before minor disturbances escalate into full-scale crises. They allow policymakers to design proactive strategies that maintain financial order, protect employment, and secure sustainable growth, especially in an interconnected global economy where shocks in one region can quickly spread across borders.

They can also help financial institutions strengthen risk management and improve resilience to external shocks. With better predictive insights, banks, insurers, and investment firms can allocate capital more prudently, design hedging strategies, and ensure liquidity even under stress. At the same time, these systems offer investors and the public a more transparent information environment, reducing uncertainty and slowing the spread of irrational panic. Improved transparency promotes confidence, stabilizes markets, and fosters healthier financial behavior. Thus, exploring intelligent models with spatiotemporal awareness is both a frontier of research and an urgent requirement for practice [6].

In summary, introducing spatiotemporal Transformers into financial crisis early warning is both a natural outcome of technological progress and an effective response to real challenges [7]. This approach enables a more comprehensive integration of temporal evolution and spatial diffusion characteristics, allowing the detection of potential crisis signals in complex financial ecosystems at an earlier stage. Advancing research in this direction is expected to promote the intelligent upgrading of financial risk governance systems, enhance resilience to systemic financial risks at both national and societal levels, and provide strong technical support for global financial stability and sustainable development.

II. Proposed Approach

From a methodological perspective, this study adopts the spatiotemporal Transformer as the central modeling framework to address the joint challenges of temporal dependence and spatial diffusion in financial risk analysis. Recent advances in causal reasoning, representation learning, and robust time-series modeling guide the design of the proposed model.

Inspired by Ying et al. [8], who demonstrated that AI-based causal reasoning over knowledge graphs can reveal hidden dependency structures and support intervention-oriented analysis, we apply attention-based mechanisms to model latent causal relationships among financial indicators across time and regions. This enables the framework to capture not only statistical correlations but also potential transmission pathways of financial risk. Following the work of Li et al. [9], who showed that causal representation learning can improve robustness and interpretability in audit risk

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2026 doi:10.20944/preprints202603.1285.v1

identification, we adopt representation learning strategies that emphasize disentanglement and stability. In our framework, embedding layers and multi-head self-attention are employed to learn reliable latent features that remain consistent under market volatility and structural changes. Moreover, Shu et al. [10] proposed a self-supervised learning framework for handling imbalanced and heterogeneous time-series data in anomaly detection. Motivated by their findings, we incorporate global feature interaction mechanisms to enhance the model's sensitivity to rare but critical risk signals, while reducing performance degradation caused by data imbalance and noise.

Based on these methodological insights, the proposed spatiotemporal Transformer integrates temporal encoding, spatial adjacency modeling, and attention-based representation learning into a unified framework. This design enables effective modeling of dynamic risk propagation and inter-regional spillover effects, and can be formally expressed as:

$$X = \{x_1, x_2, \ldots, x_T\}, \qquad x_t \in R^d \tag{1}$$

where T represents the time length and d represents the feature dimension at each moment. To simultaneously model the financial interactions between regions, a spatial adjacency matrix $A \in R^{N \times N}$ is introduced, where N is the number of different markets or regions. By introducing temporal encoding and spatial constraints in the embedding layer, the initial representation can be obtained:

$$H_0 = f_{embed}(X, A) \tag{2}$$

This representation serves as the input of the subsequent attention mechanism and provides the basis for spatiotemporal feature fusion. The overall model architecture is shown in Figure 1.

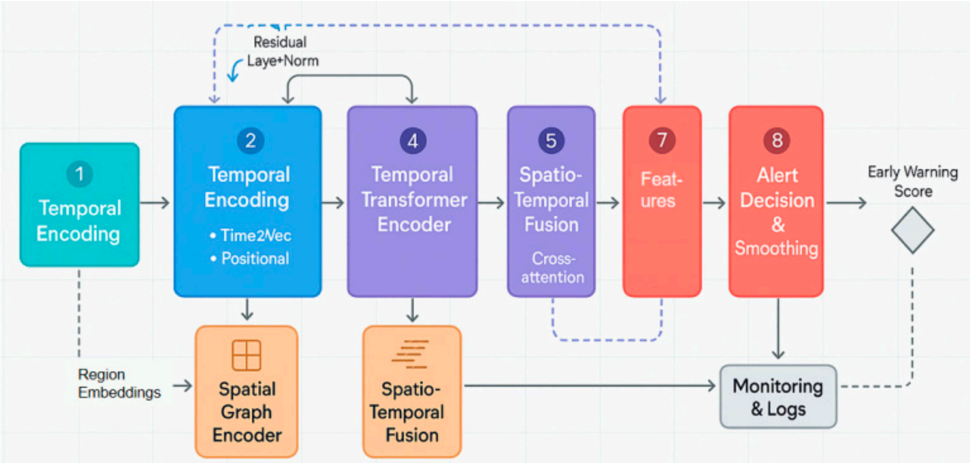

Figure 1. Overall model architecture.

In the process of sequence modeling, the core of the Transformer is the self-attention mechanism. For the input representation A, it is first mapped into the query matrix Q, the key matrix K, and the value matrix V. The calculation method is:

$$Q = HW_Q, \quad K = HW_K, \quad V = HW_V \tag{3}$$

where W_Q, W_K, W_V is the learnable parameter matrix. The attention weight is then calculated by the dot product:

 doi:10.20944/preprints202603.1285.v1

$$Attention(Q,K,V) = \mathrm{softmax}(\frac{QK^T}{\sqrt{d_k}})V \quad (4)$$

This mechanism is designed to apply self-attention for capturing long-range temporal dependencies, thus enabling global modeling of the evolving nature of financial risks. To achieve this, we adopt the transformer modeling technique [11], which has proven effective for extracting complex temporal features from heterogeneous data. By transferring this technical paradigm to the financial domain, our framework more accurately discerns latent risk signals amidst noisy time series.

Additionally, our design leverages the transaction graph-integrated transformer architecture [12], which explicitly models interactions between entities in networked transaction data for anti-money laundering applications. Integrating this methodology allows our model to account for spatial or relational dependencies among financial institutions and markets, capturing the propagation of risk across regions and sectors. Further, we apply the deep attention-based systemic risk forecasting methodology [13], which demonstrates that multi-head attention can model both short-term volatility and long-term dependencies in financial time series. Incorporating this method into our framework ensures sensitivity to both local anomalies and global market trends, ultimately enhancing the timeliness and accuracy of early financial risk warnings.

Through the integration of these advanced attention-based modeling techniques, our framework achieves robust global feature interaction and comprehensive risk signal extraction in complex, dynamic financial systems.In the spatial dimension, considering the complexity of risk transmission between markets, a graph convolution mechanism is introduced to integrate adjacency relationships. Let the nodes of a certain layer be represented by $H^{(l)}$, then the update rule is:

$$H^{(l+1)} = \sigma(\tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}H^{(l)}W^{(l)}) \quad (5)$$

where $\tilde{A} = A + I$ represents the adjacency matrix after adding a self-connection, $\tilde{D}$ is the corresponding degree matrix, $W^{(l)}$ is the learnable parameter, and $\sigma(\cdot)$ is the nonlinear activation function. Through this process, temporal dependency modeling and spatial relationship encoding can be simultaneously implemented in the spatiotemporal Transformer, thereby comprehensively characterizing the cross-regional diffusion of risk signals.

Finally, based on the obtained spatiotemporal fusion representation, we integrate a feedforward neural network and normalization layers to construct a unified risk feature representation, which serves as the basis for downstream financial risk prediction.

Specifically, the Causally Constrained Representation Learning methodology [14] is applied to ensure that the risk features are both robust to spurious correlations and interpretable with respect to causal structures in financial systems. By embedding causal graph constraints into the representation process, the learned features better capture the genuine dependencies among financial indicators.

To further enhance the adaptability and generalization ability of the model, the Dynamic Graph Deep Framework [15] is leveraged. This approach enables the architecture to dynamically model evolving risk relationships and capture the propagation of financial distress across interconnected entities and regions.

For efficient training and improved convergence in heterogeneous financial environments, the Deep Q-Learning-based Optimization strategy [16] is adopted to optimize the feedforward and normalization components, enhancing model stability and learning efficiency.

Finally, feature extraction is refined by integrating Multi-Head Self-Attention Anomaly Identification [17], which allows the network to focus on critical, subtle anomalies and sudden changes in transactional patterns, thereby improving the early detection of emerging risks.

doi:10.20944/preprints202603.1285.v1

Let the high-order spatiotemporal representation after fusion be denoted as Z; then the final prediction function is defined as:

$$y_t = f_{output}(Z_t) \tag{6}$$

where y_t represents the potential risk level at time t. The optimization goal of the entire model is to minimize the difference between the predicted value and the actual risk signal, usually in the form of:

$$L = \frac{1}{T}\sum_{t=1}^{T} l(\widehat{y}_t, y_t) \tag{7}$$

where $l(\cdot)$ is the loss function, and $\widehat{y}_t$ is the true risk marker. This optimization process ensures that the model can learn a stable and forward-looking risk representation under the constraints of spatiotemporal interactions, providing a solid algorithmic foundation for early warning of financial crises.

III. Performance Evaluation

A. Dataset

In this study, the data are derived from the Global Financial Database (GFD), which provides a multidimensional set of financial indicators. The database covers macroeconomic and financial market data of major economies over long time spans. It includes stock indices, interest rates, exchange rates, bond yields, and key indicators related to international trade. The dataset has broad coverage and a complete time series, offering a solid foundation for the study of the spatiotemporal evolution of systemic risk.

A key feature of this dataset is its multidimensional heterogeneity. In addition to common market price and return indicators, it also contains macro variables closely related to financial stability, such as inflation rate, GDP growth rate, and total credit volume. These cross-market and cross-regional data allow the study to capture not only internal fluctuations within a single market but also the linkages between different markets and the effects of risk spillovers.

At the same time, the high-frequency nature of the dataset provides favorable conditions for early warning research. By using daily and monthly indicators, it is possible to observe risk signals across different time scales. When combined with cross-country comparisons on the spatial dimension, the dataset helps reveal potential pathways of financial risk spreading from local markets to the global level. This combination of temporal and spatial features makes the dataset suitable not only for traditional economic research but also for intelligent modeling based on spatiotemporal Transformers, providing rich experimental resources and application scenarios.

B. Experimental Results

This paper first conducts a comparative experiment, and the experimental results are shown in Table 1.

Table 1. Comparative experimental results.

Method	Acc	Precision	Recall	F1-Score
ResNet	0.874	0.861	0.853	0.857
Transformer	0.892	0.881	0.873	0.877
LSTM	0.884	0.872	0.868	0.870
LSTM+CNN	0.901	0.889	0.883	0.886

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2026 doi:10.20944/preprints202603.1285.v1

OURS	0.928	0.917	0.912	0.914

From the table results, it can be seen that traditional convolutional neural networks show limited performance in financial time series risk modeling [18–21]. Although they have certain advantages in feature extraction, they cannot capture temporal dependence and cross-regional dynamics. As a result, their accuracy and F1-Score remain at relatively low levels. This indicates that relying only on spatial features is insufficient to capture the complex evolution of financial market risk signals.

In contrast, Transformer and LSTM demonstrate stronger capabilities in modeling temporal features. Transformer, through its global attention mechanism, effectively captures long-term dependencies. LSTM shows advantages in handling local time series patterns. The improved performance of both models confirms the importance of temporal dependence in the early warning of financial crises. However, these models still have limitations in modeling spatial diffusion and cross-regional risk transmission, which prevent them from achieving optimal results.

When a combined LSTM+CNN structure is introduced, model performance improves further. The convolutional layer enhances feature representation by capturing local patterns. Together with the long-term dependency modeling of LSTM, the model establishes a closer connection between temporal and local spatial features. This fusion brings notable improvements in Recall and Precision, indicating greater robustness in capturing financial risk fluctuations and abnormal signals.

Finally, the proposed model in this study achieves the best performance across all metrics. By adopting a unified spatiotemporal Transformer framework, the model not only captures long-term temporal dependencies but also integrates spatial adjacency relations. It characterizes risk spillover effects across regions more effectively. This deep integration of temporal and spatial features enables stronger early warning ability under complex financial environments and provides more reliable and forward-looking technical support for early detection of financial crises.

This paper also gives an evaluation of the model's robustness under different data missing rates, and the experimental results are shown in Figure 2.

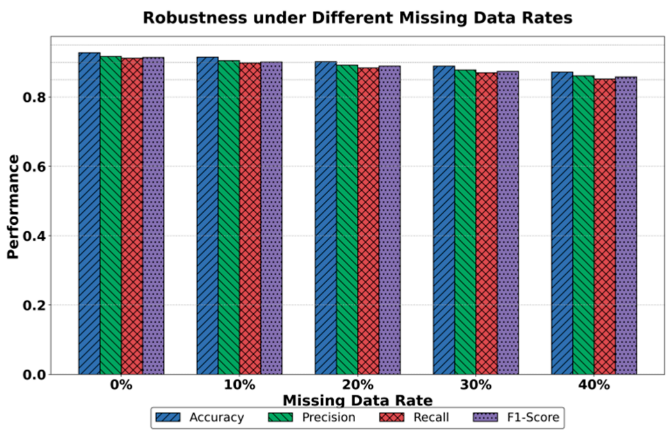

Figure 2. Model robustness evaluation under different data missing rates.

From the figure results, it can be seen that under different missing rate conditions, the overall performance of the model shows a gradual decline. However, the decline is relatively mild, which indicates that the proposed method maintains strong robustness when facing incomplete data. Even when the missing rate reaches 40 percent, the model still sustains high performance across metrics, demonstrating the stable behavior of the spatiotemporal Transformer in complex financial data environments.

In terms of accuracy and precision, the values decrease slightly as the missing rate increases, but the overall trend remains consistent. This suggests that the model has redundancy and tolerance when capturing key risk features. Since financial data often contains imbalances and missing values, these results show that the model can still extract critical features under information loss, thereby maintaining stable early warning ability.

Recall and F1-Score show a slightly greater decline compared with other metrics. This implies that under conditions of severe data loss, the sensitivity of the model to certain potential risk signals is reduced. Nevertheless, the overall performance remains above the expected level of traditional methods. This indicates that incorporating spatial adjacency relations and attention mechanisms helps to mitigate the negative impact of missing information on the model.

Overall, these experimental results confirm the applicability and reliability of the model in incomplete data environments. This is of great significance for early warning of financial crises, since data loss and uncertainty are inevitable in real scenarios. The ability of the model to remain stable under missing conditions further enhances its potential and value for application in real financial systems.

This paper further presents a sensitivity analysis of different time window lengths on early warning performance, and the experimental results are shown in Figure 3.

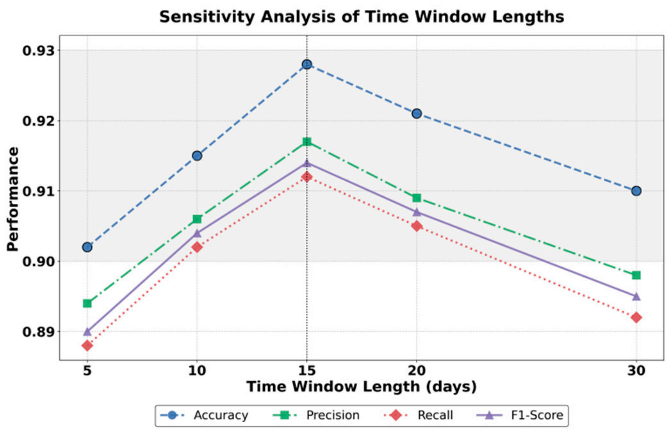

Figure 3. Sensitivity analysis of different time window lengths on early warning performance.

From the figure results, it can be observed that the model shows clear differences under different time window lengths. With a 15-day window, all indicators reach their peak, and the performance is the best. This indicates that under a medium time span, the model can better balance short-term fluctuations and long-term trends, thus capturing the most representative risk signals. This result confirms that the choice of time window plays a critical role in early warning of financial crises.

When the time window is too short, such as 5 days or 10 days, the performance of the model declines significantly. The reason is that such a short range cannot fully reflect the accumulation process of market risk. The features captured by the model are biased toward local fluctuations, while deeper trend information is ignored. Therefore, although a short window responds quickly, it struggles to build a stable risk identification ability. When the time window is too long, such as 20 days or 30 days, the model performance also decreases. This is because an overly long span introduces excessive noise and irrelevant information. As a result, the precision of distinguishing real risk signals from normal market fluctuations is reduced. The evolution of financial risk often involves nonlinearity and sudden changes, and a long window may dilute key risk patterns, thereby weakening the early warning effect.

Overall, the experimental results show that a moderate time window length is crucial for model performance. In this study, the 15-day window provides the optimal balance point, enabling the model to capture long-term dependencies while maintaining prediction sensitivity. This offers valuable guidance for the practical application of early warning models for financial crises. It helps to set reasonable time scales in real scenarios, thus achieving more accurate and robust risk monitoring.

IV. Conclusion

This study addresses the problem of early warning of financial crises and proposes a modeling framework based on spatiotemporal Transformers. By considering both temporal dependence and spatial relationships, the model can capture the dynamic evolution of financial market risks across multiple dimensions. This enables the early identification of potential crisis signals in complex environments. The results show that traditional methods often model only time series or local spatial relations in isolation. The approach presented here overcomes this limitation by highlighting the key role of multidimensional feature integration in financial risk analysis and providing new perspectives and methods for related research.

At the application level, this study not only enriches the toolbox of financial risk modeling but also provides theoretical support for regulators and financial institutions in building more intelligent risk prevention systems. With more refined risk perception and early warning mechanisms, it is possible to mitigate market panic and reduce the impact of potential crises on the real economy. This method has important implications for enhancing the stability and resilience of financial systems. In the context of globalization, where risk transmission between markets has become increasingly significant, an integrated spatiotemporal early warning model is well-suited to meet practical needs.

In addition, the proposed framework is not limited to the financial domain. It can be extended to other risk prediction scenarios where spatiotemporal dependence exists. Examples include public health, energy scheduling, and supply chain management, where risk factors also display temporal evolution and spatial diffusion. With this framework, higher prediction accuracy and stronger robustness can be achieved in cross-domain applications, further advancing the development of intelligent decision-support systems. Thus, the impact of this study lies not only in financial crisis warning but also in its cross-disciplinary applicability and methodological value.

Overall, the contribution of this study lies in the unified spatiotemporal Transformer modeling that enhances both the comprehensiveness and the forward-looking nature of risk identification. It provides solid technical support for early warning systems of financial risk. The findings hold high practical value and strong potential for application, offering new solutions for financial supervision and market participants. At the same time, they lay the foundation for risk prediction in complex systems across multiple domains. With continued technological advancement, this direction is expected to exert a lasting influence at both theoretical and practical levels and drive the evolution of risk management methodologies.

References

1. Chen, S., Huang, Y., and Ge, L., "An early warning system for financial crises: A temporal convolutional network approach," Technological and Economic Development of Economy, vol. 30, no. 3, pp. 688-711, 2024.
2. Su, T., Li, R., Liu, B., et al., "Anomaly detection and risk early warning system for financial time series based on the WaveLST-Trans model," 2025.
3. Boyle, D., and Kalita, J., "Spatiotemporal transformer for stock movement prediction," arXiv:2305.03835, 2023.
4. Sun, W., Xu, Z., Zhang, W., et al., "Advanced risk prediction and stability assessment of banks using time series transformer models," Proceedings of the 2024 5th International Conference on Big Data Economy and Information Management, pp. 657-661, 2024.
5. Namaki, A., Eyvazloo, R., and Ramtinnia, S., "A systematic review of early warning systems in finance," arXiv:2310.00490, 2023.
6. Wu, Y., Xie, L., Wan, H., et al., "A pretrained spatio-temporal hypergraph transformer for multi-stock trend forecasting," Mathematics, vol. 13, no. 10, p. 1565, 2025.
7. Sefidi, M. K., "Predicting financial market crises using multilayer network analysis and LSTM-based forecasting of spillover effects," arXiv:2505.11019, 2025.

 doi:10.20944/preprints202603.1285.v1

8. Ying, R., Liu, Q., Wang, Y., and Xiao, Y., "AI-based causal reasoning over knowledge graphs for data-driven and intervention-oriented enterprise performance analysis," 2025.
9. Li, J., Gan, Q., Wu, R., Chen, C., Fang, R., and Lai, J., "Causal representation learning for robust and interpretable audit risk identification in financial systems," 2025.
10. Shu, Y., Zhou, K., Ou, Y., Yan, R., and Huang, S., "A self-supervised learning framework for robust anomaly detection in imbalanced and heterogeneous time-series data," 2025.
11. Xie, A., and Chang, W. C., "Deep learning approach for clinical risk identification using transformer modeling of heterogeneous EHR data," arXiv:2511.04158, 2025.
12. Wu, Y., Qin, Y., Su, X., and Lin, Y., "Transformer-based risk monitoring for anti-money laundering with transaction graph integration," Proceedings of the 2025 2nd International Conference on Digital Economy, Blockchain and Artificial Intelligence, pp. 388-393, 2025.
13. Xu, Q., Xu, W., Su, X., Ma, K., Sun, W., and Qin, Y., "Enhancing systemic risk forecasting with deep attention models in financial time series," Proceedings of the 2025 2nd International Conference on Digital Economy, Blockchain and Artificial Intelligence, pp. 340-344, 2025.
14. Lai, J., Chen, C., Li, J., and Gan, Q., "Explainable intelligent audit risk assessment with causal graph modeling and causally constrained representation learning," 2025.
15. Chiang, C. F., Li, D., Ying, R., Wang, Y., Gan, Q., and Li, J., "Deep learning-based dynamic graph framework for robust corporate financial health risk prediction," 2025.
16. Gao, K., Hu, Y., Nie, C., and Li, W., "Deep Q-learning-based intelligent scheduling for ETL optimization in heterogeneous data environments," arXiv:2512.13060, 2025.
17. Wang, Y., Fang, R., Xie, A., Feng, H., and Lai, J., "Dynamic anomaly identification in accounting transactions via multi-head self-attention networks," arXiv:2511.12122, 2025.
18. He, K., Zhang, X., Ren, S., et al., "Deep residual learning for image recognition," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778, 2016.
19. Vaswani, A., Shazeer, N., Parmar, N., et al., "Attention is all you need," Advances in Neural Information Processing Systems, vol. 30, 2017.
20. Hochreiter, S., and Schmidhuber, J., "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997.
21. Jeong, D. U., and Lim, K. M., "Combined deep CNN–LSTM network-based multitasking learning architecture for noninvasive continuous blood pressure estimation using difference in ECG-PPG features," Scientific Reports, vol. 11, no. 1, p. 13539, 2021.