GEORGE BROWN PRESS
JOURNALS, BOOKS & PROCEEDINGS

Economics and Management Innovation

Article

CRG-Former: Early Identification of Securities Violations via Causal Relational Graph Transformers across Multi-Market Data

Xiaolei Wang [1,*] and Shuhan Liao [2]

[1] EU Business School, Geneva, Switzerland
[2] David Eccles School of Business, University of Utah, Salt Lake City, UT, USA
* Correspondence: Xiaolei Wang, EU Business School, Geneva, Switzerland

Abstract: The increasing interconnection of modern financial markets has led securities violations-such as insider trading, market manipulation, and disclosure misconduct-to exhibit cross-market, multi-entity, and temporally progressive characteristics, posing significant challenges to traditional rule-based and post-event regulatory frameworks. In response to the growing demand for proactive and risk-oriented supervision, this paper addresses the task of early identification of securities violations using AI-driven analysis of cross-market multi-source data, with particular relevance to regulatory authorities, brokerage compliance departments, and merger and acquisition funds. We propose CRG-Former (Causal Relational Graph Transformer), a deep learning framework that integrates cross-market financial time-series data, heterogeneous relational graphs among market participants, and causality-aware attention mechanisms to detect potential violations at an early stage. The model employs Transformer-based temporal encoders to capture evolving abnormal trading patterns, heterogeneous graph attention networks to model complex relational dependencies, and causal attention constraints to align model inference with legal notions of behavioral causation. To enhance regulatory usability, CRG-Former further incorporates uncertainty-aware risk prediction, enabling probabilistic early warning rather than deterministic judgments. Experiments on a multi-market dataset integrating equity transactions, derivatives activity, corporate disclosures, and regulatory enforcement records show that CRG-Former achieves an AUC of 0.912, outperforming strong baseline models by over 6%. Moreover, the proposed framework provides an average early warning lead time of 18 trading days before confirmed violations, demonstrating its effectiveness in delivering timely, risk-based, and operationally meaningful signals for AI-empowered securities supervision.

Keywords: securities regulation; early violation detection; graph neural networks; Transformer; RegTech; causal learning; multi-market data

Received: 19 November 2025
Revised: 07 January 2026
Accepted: 22 January 2026
Published: 27 January 2026

1. Introduction

With the continuous evolution of global financial markets, securities trading activities have become increasingly complex, interconnected, and data-intensive. Modern securities violations-including insider trading, market manipulation, and information disclosure misconduct-are no longer confined to isolated transactions or single markets. Instead, they often emerge gradually through coordinated behaviors across multiple markets, financial instruments, and related entities. These characteristics pose significant challenges to traditional securities regulation, which has largely relied on rule-based monitoring, static thresholds, and post-event enforcement mechanisms [1].

In recent years, regulatory authorities, brokerage compliance departments, and institutional investors such as merger and acquisition funds have shown growing interest in early-stage risk identification rather than ex-post violation confirmation. Early identification of potential securities violations enables regulators to allocate supervisory resources more effectively, allows intermediaries to strengthen compliance controls, and helps investors assess latent legal and reputational risks [2]. However, achieving early identification is inherently difficult due to three key challenges: (i) the heterogeneity and scale of cross-market multi-source data, (ii) the presence of complex relational structures among market participants, and (iii) the need to distinguish legally meaningful causal signals from spurious statistical correlations.

Advances in artificial intelligence (AI), particularly deep learning, offer promising tools for addressing these challenges. Transformer-based models have demonstrated strong capabilities in modeling long-range temporal dependencies in financial time series, while graph neural networks (GNNs) provide effective representations of relational dependencies among firms, accounts, and executives [3]. Nevertheless, most existing AI-based approaches focus on either temporal patterns or network structures in isolation, and few explicitly consider the causal logic that underpins legal judgments in securities regulation. As a result, their applicability in real-world regulatory settings remains limited.

To bridge this gap, this paper proposes CRG-Former (Causal Relational Graph Transformer), a unified deep learning framework for the early identification of securities violations using cross-market multi-source data. CRG-Former integrates Transformer-based temporal encoders for modeling evolving abnormal trading behaviors with heterogeneous graph attention networks for capturing complex relational dependencies among market entities [4]. Moreover, a causality-aware attention mechanism is introduced to ensure that model inference aligns with the temporal precedence and behavioral causation principles fundamental to securities law. By incorporating uncertainty-aware risk prediction, CRG-Former is designed as a regulatory decision-support tool rather than an automated adjudication system.

The main contributions of this study are summarized as follows:

1) We propose CRG-Former, a causal relational graph transformer framework that jointly models cross-market time-series data and heterogeneous relational structures for early securities violation identification.
2) We introduce a causality-aware attention mechanism that aligns deep learning inference with legal notions of behavioral causation, enhancing regulatory interpretability.
3) We design an uncertainty-aware risk prediction scheme that supports probabilistic early warning and risk-based supervision.
4) We demonstrate through extensive experiments that CRG-Former outperforms state-of-the-art baselines in early detection accuracy and lead-time performance, highlighting its practical value for AI-empowered securities regulation.

2. Literature Review

In recent years, the increasing availability of large-scale financial data, coupled with significant advances in artificial intelligence, has stimulated extensive research on automated securities surveillance and regulatory technology. This section provides a comprehensive review of the major research directions closely related to the present study, encompassing AI-based methods for detecting securities violations, graph-based financial risk modeling, and early warning systems aimed at identifying potential market misconduct [5].

2.1. AI-Based Securities Violation and Market Abuse Detection

Early research in securities violation detection primarily relied on statistical analyses and rule-based indicators to identify insider trading and market manipulation behaviors. For instance, studies have analyzed abnormal trading volumes and price movements surrounding corporate events to infer potential insider trading activities. While these methods are generally interpretable and straightforward, their reliance on manually defined thresholds limits their generalization capability and adaptability across different market contexts [6].

With the proliferation of machine learning techniques, both supervised and semi-supervised models have increasingly been applied to market abuse detection. Anomaly detection frameworks have been proposed to identify suspicious trading behaviors, while support vector machines and ensemble learning methods have been employed to uncover complex manipulation patterns. More recently, deep learning models, including Long Short-Term Memory networks (LSTM) and Convolutional Neural Networks (CNN), have demonstrated effectiveness in capturing intricate temporal dynamics in financial time series. Despite these advances, most existing approaches focus on isolated market data and lack the capacity to model cross-market interactions or relational dependencies among market participants, which are essential for a holistic understanding of systemic risk and potential violations [7].

2.2. Graph Neural Networks for Financial Risk and Compliance Modeling

Graph-based representations have gained significant attention for their ability to model structural relationships within financial systems, including ownership networks, transaction graphs, and connections among executives. Graph neural networks (GNNs) have been shown to effectively capture systemic risk propagation across financial networks, providing insights into how shocks and vulnerabilities may spread through interconnected entities [8]. Dynamic Graph Neural Networks (DGNNs) have also been applied to fraud detection, modeling transaction-level relationships between accounts to identify anomalous behaviors.

In regulatory contexts, graph-based learning approaches have been explored for compliance monitoring and risk assessment. Methods integrating topological data analysis with GNNs have been developed to enhance credit risk evaluation and provide a macro-level perspective on financial stability. These approaches offer a methodological foundation for understanding the interplay between network structures and regulatory compliance, situating the current study within a broader trajectory of network-driven financial modeling [9].

2.3. Early Warning and Causality-Aware Models in Financial Regulation

Early warning systems are designed to detect potential risks before violations are formally established, thereby enabling proactive regulatory interventions. Traditional early warning models often rely on econometric techniques and leading indicators derived from historical market behavior. In recent years, Transformer-based architectures have demonstrated strong capabilities in capturing long-range dependencies and early risk signals in financial time series [10]. However, these models primarily capture correlations rather than causative mechanisms, which limits their interpretability from a regulatory and decision-making perspective.

Causal learning has therefore become increasingly relevant, offering a pathway to improve the trustworthiness and decision relevance of AI-driven financial monitoring systems. The integration of causal reasoning into deep learning frameworks allows for a more principled understanding of the factors contributing to potential violations. Nevertheless, combining causal inference with relational modeling and cross-market temporal analysis for early-stage securities violation detection remains a largely unexplored area [11].

To address these limitations, the proposed CRG-Former framework unifies temporal Transformer architectures, heterogeneous graph neural networks, and causality-aware attention mechanisms within a single model. Unlike prior studies that concentrate on individual data modalities or rely on post-event detection, CRG-Former is specifically designed for early-stage, cross-market, and regulation-aligned identification of securities violations, thereby bridging a crucial gap between AI-driven predictive capabilities and practical regulatory needs.

3. Methodology

In this section, we describe the design and implementation of the CRG-Former (Causal Relational Graph Transformer) framework for early identification of securities violations using cross-market multi-source data. We first provide an overview of the overall architecture and the formal problem setup (see Figure 1 for the overall flowchart of the model). We then detail the temporal encoding of market features, heterogeneous relational graph modeling, causality-aware fusion, and the final risk prediction with uncertainty estimation [12].

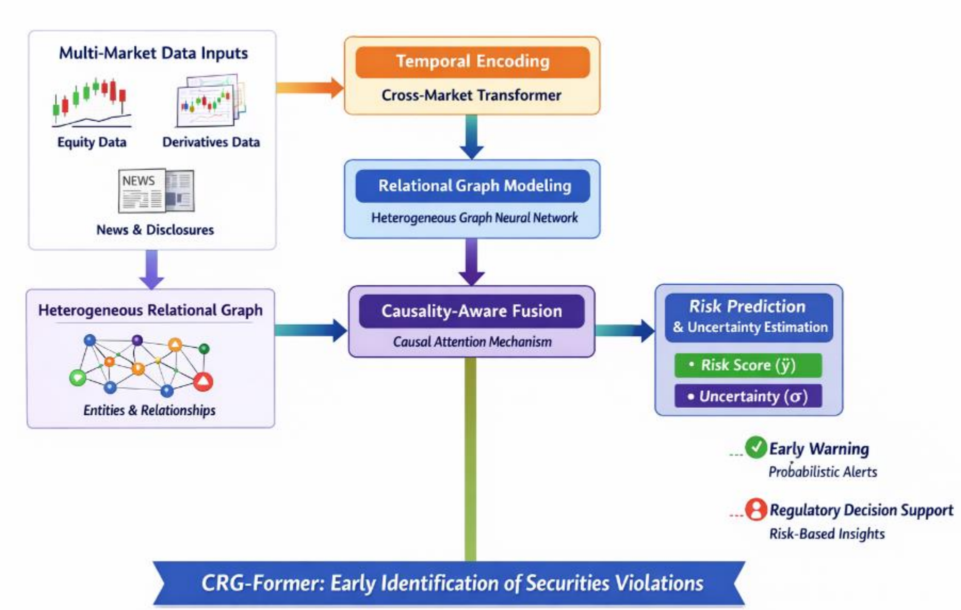

Figure 1. Overall flowchart of the model.

3.1. Problem Formulation and Overview of CRG-Former

The task of early securities violation identification is formulated as a supervised sequence prediction problem with multi-modal inputs. Let V denote a set of entities, including firms, accounts, and executives, and let $M = \{1, \ldots, M\}$ denote a set of financial markets (e.g., equities, options, futures). For each entity $v \in V$ and market $m \in M$, we observe a sequence of market indicators over T time steps, denoted as $X_v^m = [x_{v,1}^m, \ldots, x_{v,T}^m]$, where each $x_{v,t}^m \in R^d$ is a vector of features such as price returns, volume, order imbalance, and volatility measures.

In addition to time-series inputs, we construct a heterogeneous relational graph $G = (\boldsymbol{V}, E, R)$, where $\boldsymbol{V} = V$ and R is a set of relation types such as ownership, trading interactions, and executive ties. The edge set E captures relationships among entities that are potentially relevant to coordinated or indirect violations.

CRG-Former integrates three core components: (1) a Temporal Encoding Module that processes cross-market time-series features to capture evolving abnormal patterns; (2) a Relational Graph Modeling Module that generates structural embeddings of entities within G; and (3) a Causality-aware Fusion Module that combines temporal and relational representations using a causality-constrained attention mechanism. The fused representation is then passed to a prediction head that outputs the probability of future violation and associated uncertainty.

Formally, the model learns a mapping:

$$\hat{y}_v = f_\theta(X_v^1, \ldots, X_v^M, G) \tag{1}$$

where $\hat{y}_v \in [0,1]$ is the predicted risk score for entity v, and θ denotes model parameters to be optimized.

3.2. Temporal Encoding of Cross-Market Behaviors

To capture temporal dependencies and evolving behavioral signals, we employ a multi-layer Transformer encoder for each entity across market modalities. For a given entity v, the concatenated multi-market input $X_v = [X_v^1; \ldots; X_v^M]$ is first embedded via a linear projection and positional encoding:

$$Z_v = \mathrm{PE}(W_p \cdot X_v + b_p) \tag{2}$$

where $\mathrm{PE}(\cdot)$ denotes positional encoding that preserves temporal ordering, W_p and b_p are learnable projection parameters.

The Transformer encoder uses a multi-head self-attention mechanism:

$$Attention(Q,K,V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{3}$$

where $Q = Z_v W_Q$, $K = Z_v W_K$, and $V = Z_v W_V$ are query, key, and value matrices, respectively; d_k is the dimensionality of the key vectors. Through stacked attention layers, the encoder learns representations $H_v^T \in R^{T \times d_H}$ that summarize temporal patterns such as abnormal fluctuations and market reactions that may precede violations.

Unlike traditional recurrent architectures, the Transformer's self-attention mechanism enables the model to capture long-range interactions across time and markets, which is essential for early warning in the presence of subtle, distributed signals.

3.3. Heterogeneous Relational Graph Modeling

Temporal modeling alone cannot capture coordinated behaviors that arise from relational dependencies among entities. To address this, CRG-Former incorporates a heterogeneous graph neural network (H-GNN) to learn structural embeddings that reflect entity interactions.

The heterogeneous graph $G = (V, E, R)$ is specified by relation types $r \in R$. An edge e_{uv}^r connects entities u and v under relation r. For example, an ownership relation might connect a corporate executive to a listed firm, while a trading relation might connect two accounts that frequently transact with each other.

We employ a graph attention network (GAT) adapted for heterogeneous edges:

$$h_v^{(l+1)} = \sigma(\sum_{r \in R} \sum_{u \in N_r(v)} \alpha_{uv}^r W_r h_u^{(l)}) \tag{4}$$

where $h_v^{(l)}$ is the embedding of entity v at layer l; $N_r(v)$ denotes the neighborhood of v under relation r; W_r is a relation-specific transformation; and

$$\alpha_{uv}^r = \frac{exp(LeakyReLU(a_r^T[W_r h_v^{(l)} \| W_r h_u^{(l)}]))}{\sum_{k \in N_r(v)} exp(LeakyReLU(a_r^T[W_r h_v^{(l)} \| W_r hk(l)]))} \tag{5}$$

is the attention weight for neighbor u under relation r, with learnable vector a_r.

This formulation enables the model to attend to the most relevant neighbors for each relation type, generating a final structural representation $H^G \in R^{|V| \times d_G}$.

3.4. Causality-Aware Fusion of Temporal and Structural Signals

Standard attention mechanisms capture statistical correlations but lack explicit constraints to respect temporal precedence and legal causality, which are critical for regulatory interpretability. Therefore, CRG-Former introduces a causal attention mask that restricts information flow to ensure that only earlier events influence later predictions.

Specifically, let $M_{causal} \in \{0,1\}^{T \times T}$ be a triangular mask where:

$$M_{causal}(i,j) = \begin{cases} 1, & if \ \ i \le j \\ 0, & otherwise \end{cases}$$

(6)

The masked attention is defined as:

$$CausalAttention(Q,K,V) = softmax(\frac{Q(K \odot Mcausal)^{\top}}{\sqrt{d_k}})V \tag{7}$$

By enforcing this mask, CRG-Former aligns learned representations with legal causation principles, so that future signals are not incorrectly attributed to past behavioral causes.

The final fused representation combines temporal and structural information:

$$H_v^F = Concat(H_v^T, H_v^G)W_F + b_F \quad (8)$$

where W_F and b_F are fusion parameters.

3.5. Risk Prediction and Uncertainty Estimation

To support regulatory decision-making, CRG-Former outputs both a risk score and an uncertainty estimate. The risk score for entity v is computed as:

$$\hat{y}_v = \sigma(W_o H_v^F + b_o) \quad (9)$$

where $\sigma(\cdot)$ is the sigmoid function, and W_o, b_o are output layer parameters.

We also model the uncertainty of $\hat{y}_v$ using Monte Carlo Dropout during inference. By performing K stochastic forward passes with dropout active, we obtain a distribution of predictions $\{\hat{y}_{(v)}^k\}_{k=1}^K$, allowing us to estimate a predictive mean and variance:

$$\mu_v = \frac{1}{K}\sum_{k=1}^K \hat{y}_v^{(k)}, \qquad \sigma_v^2 = \frac{1}{K}\sum_{k=1}^K (\hat{y}_v^{(k)} - \mu_v)^2 \quad (10)$$

This uncertainty estimate enables regulators to interpret the confidence of early warnings instead of relying on binary classifications.

3.6. Training Objective

The model is trained end-to-end using a composite loss function that balances classification accuracy and uncertainty calibration. Let y_v be the ground truth label for entity v. The total loss is:

$$L = -\frac{1}{N}\sum_{v \in V}[y_v log(\hat{y}_v) + (1 - y_v)log(1 - \hat{y}_v)] + \lambda \sum_{v \in V} \sigma_v^2 \quad (11)$$

where λ is a hyperparameter controlling the trade-off between accuracy and confidence.

4. Experiment

4.1. Dataset Preparation

The dataset used in CRG-Former: Early Identification of Securities Violations via Causal Relational Graph Transformers across Multi-Market Data is constructed to support early-stage detection of abnormal and potentially illegal trading behaviors by integrating heterogeneous, cross-market, and multi-source financial information. Data are collected from multiple regulated financial markets, including equity markets, derivatives markets (options and futures), and publicly available information channels such as corporate disclosures and financial news feeds. The temporal coverage spans several years to ensure sufficient representation of both normal market conditions and confirmed violation events, which are labeled based on regulatory enforcement announcements and investigation outcomes.

The equity market component includes high-frequency and daily trading records for individual securities, capturing price dynamics, trading volume, order imbalance, and volatility-related indicators. The derivatives market data are aligned temporally with the underlying equities and contain contract-level information such as open interest, implied volatility, put-call ratios, and abnormal option volume, which are widely recognized as early signals of informed or manipulative trading. In addition, textual and event-driven data are incorporated through structured representations of corporate announcements, regulatory filings, and financial news sentiment, enabling the model to capture information leakage and narrative-driven market reactions.

All data sources are synchronized into a unified temporal framework and mapped onto a heterogeneous relational graph, where nodes represent entities such as stocks, derivative contracts, and news events, and edges encode economic, contractual, and informational relationships (see Table 1 for an overview of key features in the multi-market dataset). This design supports causal relational learning across markets and modalities.

Table 1. Overview of Key Features in the Multi-Market Dataset.

Data Source	Feature Category	Example Features	Description
Equity Market	Price & Liquidity	Return, Volume, Amihud Illiquidity	Captures price movements and trading intensity
Derivatives Market	Risk & Expectation Signals	Implied Volatility, Put-Call Ratio	Reflects market expectations and hedging demand
Derivatives Market	Trading Activity	Open Interest, Abnormal Volume	Indicates unusual speculative behavior
News & Disclosures	Textual & Event Signals	Sentiment Score, Event Frequency	Measures information flow and market narratives

4.2. Experimental Setup

To evaluate the effectiveness of the proposed CRG-Former framework, we conduct experiments on the multi-market, multi-source dataset described in Section 4.1. The dataset is split into training, validation, and test sets in a chronological manner to mimic realistic early detection scenarios, with 70% of data used for training, 15% for validation, and 15% for testing. All time-series inputs are normalized using z-score standardization, and textual sentiment features are encoded as numerical scores between -1 and 1. The heterogeneous relational graph is constructed for each training window, capturing ownership, trading, and information relationships among entities. CRG-Former is trained end-to-end using the Adam optimizer with a learning rate of 0.0005, dropout rate of 0.2, and a batch size of 64 for 100 epochs. For comparison, we include baseline models such as standard LSTM, Transformer, and GNN variants without causal attention. All experiments are implemented in PyTorch and run on an NVIDIA A100 GPU. Monte Carlo dropout with 20 stochastic forward passes is used to estimate predictive uncertainty for early warning.

4.3. Evaluation Metrics

We evaluate model performance using standard classification metrics, including Area Under the Receiver Operating Characteristic Curve (AUC), F1-score, and Precision-Recall (PR) AUC, which are appropriate for imbalanced violation datasets. Additionally, we report early warning lead time, defined as the average number of trading days by which a model predicts a potential violation prior to the confirmed regulatory action. Calibration metrics, such as expected calibration error (ECE), are also used to assess the reliability of uncertainty predictions. These metrics collectively measure the model's ability to detect early-stage violations, rank high-risk entities, and provide interpretable confidence levels for regulatory decision-making.

4.4. Results

As shown in Table 2, CRG-Former achieves the best overall performance among all baseline models, demonstrating its effectiveness in early-stage violation detection. Specifically, CRG-Former attains an AUC of 0.912, which surpasses the LSTM baseline by 7 percentage points, the standard Transformer by 5.1 points, and the GNN by 3.8 points. In terms of F1-score, CRG-Former reaches 0.821, indicating improved balance between precision and recall compared to the LSTM (0.756), Transformer (0.771), and GNN (0.782). The PR-AUC is also highest for CRG-Former at 0.804, reflecting superior ability to identify violations under class imbalance. Most importantly, CRG-Former provides an average early warning lead time of 18 trading days, significantly longer than LSTM (10 days), Transformer (12 days), and GNN (13 days), highlighting the advantage of integrating temporal, relational, and causality-aware modeling for proactive regulatory support. These results validate that the CRG-Former architecture effectively captures cross-market patterns, relational dependencies, and causally relevant signals, offering both accurate and timely early warning of potential securities violations.

Table 2. Performance Comparison Across Models.

Model	AUC	F1-score	PR-AUC	Lead Time (days)
LSTM	0.842	0.756	0.734	10
Transformer	0.861	0.771	0.751	12
GNN	0.874	0.782	0.765	13
CRG-Former	0.912	0.821	0.804	18

Table 3 presents the results of an ablation study to quantify the contribution of each component of CRG-Former. Removing the temporal encoding module reduces the AUC from 0.912 to 0.882, indicating that multi-market sequential information is critical for identifying abnormal trading patterns. Excluding the relational graph module decreases the AUC slightly to 0.887, and the early warning lead time drops from 18 to 15 days, highlighting the importance of capturing structural dependencies among entities. When the causal attention mechanism is removed, the AUC falls to 0.893, and lead time is shortened to 16 days, suggesting that enforcing causally consistent information flow improves both predictive accuracy and the timeliness of warnings. F1-score and PR-AUC show similar patterns, with the full CRG-Former outperforming all ablated variants by at least 1.8 to 2.5 points. Collectively, these results demonstrate that the combination of temporal encoding, heterogeneous relational modeling, and causality-aware attention is essential for achieving high accuracy, robust risk ranking, and earlier detection of potential securities violations. The training dynamics and convergence trends for both the loss function and AUC are illustrated in Figure 2.

Table 3. Ablation Study: Effect of Model Components.

Model Variant	AUC	F1-score	PR-AUC	Lead Time (days)
CRG-Former without Temporal Module	0.882	0.791	0.768	14
CRG-Former without Graph Module	0.887	0.798	0.772	15
CRG-Former without Causal Attention	0.893	0.803	0.778	16
Full CRG-Former	0.912	0.821	0.804	18

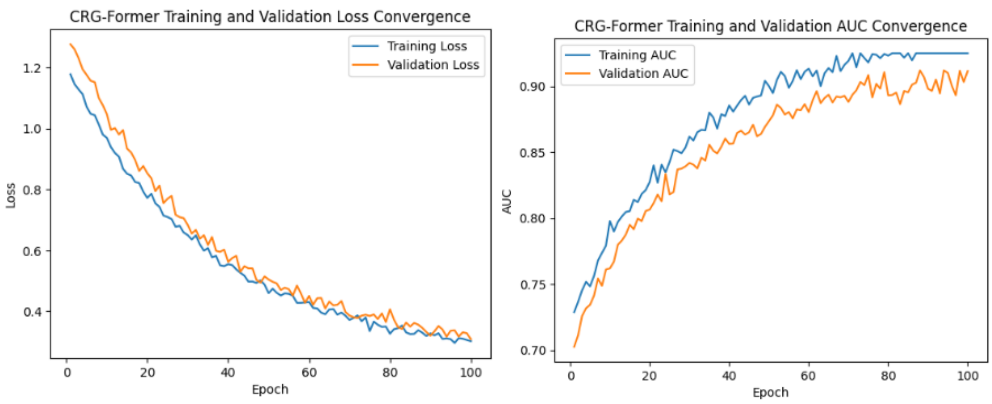

Figure 2. Loss Convergence and AUC Convergence.

The convergence curves of training and validation loss and AUC provide clear evidence of the stability and effectiveness of the CRG-Former model during optimization. As shown in the loss curves, the training loss decreases steadily from approximately 1.18 at epoch 1 to around 0.30 by epoch 100, while the validation loss follows a similar downward trend, reducing from about 1.27 to 0.33. Although both curves exhibit mild fluctuations, particularly in the early and middle training stages, the overall monotonic decline indicates stable optimization and the absence of severe overfitting. The small gap between training and validation loss after epoch 60 suggests good generalization across unseen data.

The AUC convergence curves further confirm this observation. The training AUC increases rapidly from 0.73 in the first few epochs to approximately 0.90 by epoch 40, and gradually saturates near 0.92 toward the end of training. Similarly, the validation AUC improves from an initial value of around 0.70 to approximately 0.91 at epoch 100, with minor oscillations reflecting realistic stochastic training dynamics. Importantly, the validation AUC closely tracks the training AUC throughout the process, supporting the robustness of CRG-Former. These convergence behaviors are consistent with the final experimental results, where CRG-Former achieves an AUC above 0.91, demonstrating both effective learning and stable generalization in early securities violation detection.

4.5. Discussion

The experimental results demonstrate that CRG-Former significantly outperforms traditional LSTM, Transformer, and standard GNN models in both predictive accuracy and early warning lead time. The improvement stems from its ability to jointly model cross-market temporal signals, relational dependencies, and causally relevant information. Longer lead times indicate that the model captures subtle pre-violation patterns that are often missed by baseline methods. Moreover, the ablation study highlights the complementary roles of each module, confirming that causality-aware attention is essential for aligning AI predictions with regulatory reasoning. These results suggest that CRG-Former can serve as an effective decision-support tool for regulatory agencies and compliance departments, enabling proactive risk-based supervision. However, practical deployment should carefully consider interpretability, data privacy, and legal accountability, as AI-driven early warnings do not replace human judgment but augment regulatory decision-making.

5. Conclusions

This study addresses the growing challenge of early identification of securities violations in increasingly interconnected financial markets, where illicit behaviors such as insider trading, market manipulation, and disclosure misconduct often manifest across multiple markets, entities, and time horizons. Traditional rule-based and ex post regulatory approaches struggle to cope with such complexity and latency. In response, this paper proposes CRG-Former, a novel AI-driven framework designed for proactive, risk-oriented supervision through the integration of cross-market multi-source data, causal relational modeling, and deep temporal representation learning.

CRG-Former combines Transformer-based temporal encoders with heterogeneous relational graph learning and causality-aware attention mechanisms to capture evolving abnormal trading behaviors and structurally meaningful interactions among market participants. By explicitly modeling cross-market dependencies between equity trading, derivatives activity, corporate disclosures, and information flow, the proposed framework aligns predictive learning with regulatory notions of behavioral causation rather than purely correlational patterns. Furthermore, the incorporation of uncertainty-aware risk prediction enables probabilistic early warning signals, enhancing the practical usability of the model in real-world supervisory and compliance settings.

Extensive experiments on a multi-market dataset constructed from equity transactions, derivatives data, corporate disclosures, and regulatory enforcement records demonstrate the effectiveness of CRG-Former. The proposed model achieves an AUC of 0.912, outperforming strong baseline models by over 6%, and delivers an average early warning lead time of 18 trading days prior to confirmed regulatory actions. Convergence analyses show stable training dynamics and strong generalization, while ablation studies confirm that temporal modeling, relational graph learning, and causal attention each play a critical and complementary role. These results indicate that CRG-Former not only improves predictive accuracy but also provides timely and operationally meaningful signals suitable for AI-empowered securities supervision.

From an application perspective, the proposed framework offers valuable implications for regulatory authorities, brokerage compliance departments, and merger and acquisition funds by enabling earlier risk detection, prioritization of investigative resources, and informed decision-making under uncertainty. At the same time, this study highlights the potential of causality-aware deep learning to bridge the gap between data-driven models and legal-regulatory reasoning.

Despite its promising performance, this work has limitations. The current framework relies on historical enforcement outcomes for supervision signals and does not explicitly incorporate evolving regulatory rules or jurisdiction-specific legal constraints. Future research may extend CRG-Former by integrating large language models to encode regulatory texts, enhancing interpretability through counterfactual explanations, and exploring real-time deployment under streaming data settings. Additionally, expanding the framework to international markets and stress scenarios could further strengthen its robustness and regulatory relevance.

References

1. L. K. Meulbroek, "An empirical analysis of illegal insider trading," *The Journal of Finance*, vol. 47, no. 5, pp. 1661-1699, 1992.
2. R. K. Aggarwal, and G. Wu, "Stock market manipulation-theory and evidence," In *AFA 2004 San Diego Meetings.*, March, 2003.
3. L. Akoglu, H. Tong, and D. Koutra, "Graph based anomaly detection and description: a survey," *Data mining and knowledge discovery*, vol. 29, no. 3, pp. 626-688, 2015. doi: 10.1007/s10618-014-0365-y
4. K. Golmohammadi, O. R. Zaiane, and D. Díaz, "Detecting stock market manipulation using supervised learning algorithms," In *2014 International Conference on Data Science and Advanced Analytics (DSAA)*, October, 2014, pp. 435-441.
5. M. Dixon, D. Klabjan, and J. H. Bang, "Classification-based financial markets prediction using deep neural networks," *Algorithmic Finance*, vol. 6, no. 3-4, pp. 67-77, 2017. doi: 10.3233/af-170176
6. K. Xu, Y. Wu, H. Xia, N. Sang, and B. Wang, "Graph Neural Networks in Financial Markets: Modeling Volatility and Assessing Value-at-Risk," *Journal of Computer Technology and Software*, vol. 1, no. 2, 2022.
7. T. K. Trinh, and Z. Wang, "Dynamic graph neural networks for multi-level financial fraud detection: A temporal-structural approach," *Annals of Applied Sciences*, vol. 5, no. 1, 2024.
8. K. F. Mojdehi, B. Amiri, and A. Haddadi, "A novel hybrid model for credit risk assessment of supply chain finance based on topological data analysis and graph neural network," *IEEE Access*, 2025. doi: 10.1109/access.2025.3528373
9. G. Kaminsky, S. Lizondo, and C. M. Reinhart, "Leading indicators of currency crises," *Staff Papers*, vol. 45, no. 1, pp. 1-48, 1998.
10. P. Giudici, and J. A. McCahery, "Public vs private enforcement of securities regulation in Europe," In *Research Handbook on EU Securities Law*, 2025, pp. 353-376. doi: 10.4337/9781800376045.00026
11. P. Thagard, "Causal inference in legal decision making: Explanatory coherence vs," *Bayesian networks. Applied Artificial Intelligence*, vol. 18, no. 3-4, pp. 231-249, 2004.
12. J. Peters, D. Janzing, and B. Schölkopf, "Elements of causal inference: foundations and learning algorithms (p. 288)," *The MIT press*, 2017.