Journal of Computer Technology and Software

ISSN:2998-2383

Vol. 3, No. 8, 2024

Cross-Timescale Transformer with One-Dimensional Convolution for Integrated Financial Risk Anomaly Detection and Discrimination

Yinghao Zhao
Pace University, New York, USA
yinghaozhaony@gmail.com

Abstract: This paper addresses the characteristics of risk anomalies in financial time series, such as concealed signals, multi-scale coupling, and non-stationary distributions. It proposes a risk anomaly detection and discrimination model based on one-dimensional convolution and a cross-timescale Transformer. The method takes multivariate sequences as input, employs a one-dimensional convolutional encoder to extract local fluctuation patterns and short-range mutation features, and constructs sequence representations with different time resolutions through multi-scale downsampling. Long-range dependencies and cross-cycle correlations are learned in a cross-timescale attention module. Subsequently, the multi-scale context is upsampled and aligned to a unified time axis, then fused with the local representations to form a joint representation that combines fine-grained localization capabilities with global semantic consistency. The model simultaneously outputs time-by-time anomaly scores and sequence-level risk category probabilities within the same framework, achieving integrated modeling for anomaly localization and risk discrimination. Comparative experiments show that this method achieves superior performance in classification, ranking ability, and reliability metrics, validating the effectiveness and stability of local pattern modeling and cross-scale context fusion for identifying financial risk anomalies.

Keywords: Multivariate time series analysis, outlier scoring, cross-scale attention, risk assessment

1. Introduction

Financial markets operate in a highly dynamic and uncertain environment. Risk events often manifest themselves in concealed, sudden, and cross-asset forms, exhibiting significant non-stationarity and structural changes under the combined influence of macroeconomic policies, industry structure, and investor behavior. Traditional risk identification and early warning systems rely heavily on linear assumptions, fixed distributions, or low-order statistical characteristics, making it difficult to characterize the state transitions, volatility clustering, and tail risk accumulation of price and transaction sequences at different stages, or to capture the chain reactions triggered by local disturbances during anomaly evolution[1]. Therefore, constructing models capable of stably detecting anomalies and assessing risks within complex time-series structures is not only crucial for enhancing the resilience of the financial system and the effectiveness of regulation but also places higher demands on the timeliness and interpretability of risk management[2].

Financial risk anomalies are characterized by multi-scale coupling and multi-form coexistence: at the short-term level, there are instantaneous anomalies caused by liquidity shocks, trading congestion, and sentiment fluctuations; at the medium- to long-term level, there are slow drifts and periodic revaluations resulting from fundamental changes, institutional adjustments, and structural imbalances[3]. Patterns across different time scales are often not independent but rather interconnected, forming cross-scale linkages through fluctuation propagation, correlation reconstruction, and leverage feedback. This makes single-window or single-frequency modeling prone to missed detections and false positives. Furthermore, the outward appearance of anomalies may manifest as similar local fluctuations under different market conditions, but their risk implications can be drastically different, further increasing the challenges to the consistency and robustness of anomaly detection and risk assessment[4].

In this context, the collaborative design of local pattern extraction and global dependency modeling is of great significance. One-dimensional convolution excels at efficiently representing the local structure of time series, extracting discriminative short-range fluctuation patterns and local abrupt changes from noisy backgrounds, and completing multi-level feature extraction with low computational cost. Transformers possess the ability to model long-range dependencies and global relationships, learning cross-period correlations and contextual constraints over a larger time span, thereby identifying persistent risk accumulation and cross-stage propagation triggered by local anomalies. By aligning and fusing representations at different scales through cross-timescale modeling mechanisms, it is hoped that the ability to characterize long-term structural changes and state transitions can be enhanced while maintaining local sensitivity[5].

To meet the practical needs of financial risk governance and business decision-making, anomaly detection not only needs to identify time-series segments that deviate from the norm but also requires more granular discrimination of the risk attributes of these anomalies to support risk stratification and early warning, strategy adjustment, and resource allocation. Under conditions of complex distribution drift and noise interference, models with cross-timescale understanding capabilities can better distinguish between transient disturbances and systemic risk signals, improving the stability and consistency of early warnings. Research focusing on the joint modeling approach of one-dimensional convolution and cross-timescale Transformers helps to form a unified representation framework for financial time-series anomalies, providing theoretical and methodological support for building a more reliable risk identification system and more adaptive intelligent risk control capabilities.

2. BackGround

One of the core objectives of financial risk management is to promptly identify potential risk signals in a market environment characterized by significant uncertainty and to translate them into actionable early warning and control strategies[6]. With the increasing speed of high-frequency trading, cross-market linkages, and information dissemination, the formation and spread of risk exhibit greater complexity and nonlinearity. Abnormal behavior is no longer limited to extreme values of a single indicator but often manifests as multivariate co-variable shifts, structural mutations, and cyclical rhythm disruptions. Simultaneously, financial time-series data generally suffers from high noise, missing data, and asynchronous sampling, and is significantly affected by changes in the macroeconomic environment, causing data distribution to drift across different stages. This makes it difficult for anomaly identification methods based on static thresholds or fixed pattern libraries to operate stably over the long term.

At the methodological level, early statistical tests and rule systems relied on prior settings and strong assumptions, providing intuitive interpretations of risk but with limited coverage of complex patterns. Traditional machine learning-based solutions typically rely on manual feature engineering, compressing time-series segments into fixed-dimensional representations, which are prone to declining generalization when feature design is insufficient or market structure changes occur[7]. Deep learning methods have improved feature representation capabilities through end-to-end learning, but they still face two key challenges: First, networks that emphasize local modeling are insufficient in characterizing long-term dependencies and cross-cycle relationships, making it difficult to identify slowly accumulating risks. Second, structures that emphasize global modeling may become less sensitive to noise interference and local mutations, leading to higher training and deployment costs. Therefore, for the task of detecting and identifying financial risk anomalies, an information modeling mechanism that can take into account both local details and cross-scale context is needed to adapt to the risk identification requirements in multi-form anomalies and non-stationary environments.

3. Methodology

3.1 Overall framework

This method targets anomaly detection and risk assessment for multivariate financial time series. The input is an observed sequence of length T and a set of variables with feature dimension F. Through local pattern encoding and cross-timescale modeling, it obtains time-by-time anomaly scores and sequence-level risk category distributions. The method consists of three parts: a one-dimensional convolutional local encoder to extract short-range fluctuation patterns and local mutation representations; a cross-timescale Transformer to model long-range dependencies and cross-period correlations on multi-resolution time axes, and to achieve context completion through inter-scale alignment; and a fusion and discriminator head to jointly aggregate local and cross-scale representations, outputting anomaly scores and risk category probabilities. Figure 1 shows the overall model architecture.

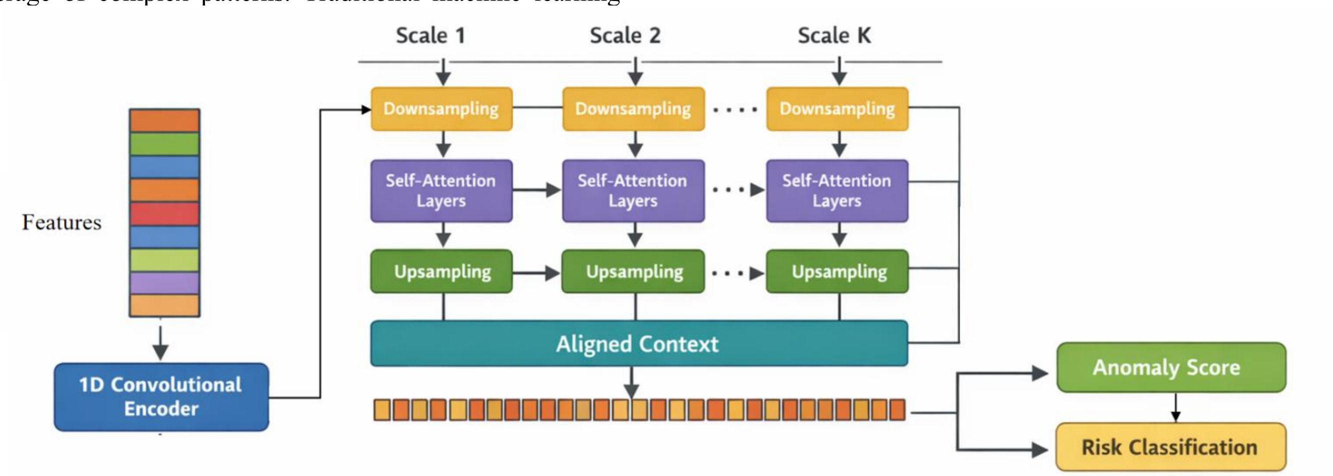

Figure 1. This figure illustrates the overall architecture of a financial risk anomaly detection and discrimination model based on one-dimensional convolution and cross-timescale Transformer, including key components such as local feature encoding, multi-scale context modeling, and time-aligned fusion. The model jointly outputs time-by-time anomaly scores and sequence-level risk category predictions on a unified time axis, achieving a collaborative characterization of local mutations and long-term dependencies.

To improve adaptability to non-stationary conditions, representation learning is based on a unified temporal embedding. It constructs multi-scale tokens using a shared parameter form but different downsampling ratios across different time scales, and then upsamples back to the original time resolution to complete the time-by-time output. The input sequence is denoted as:

$$X = [x_1, x_2, \ldots, x_T]^T \in R^{T \times F}$$

3.2 One-dimensional convolutional local encoding

The local encoder employs multi-layer one-dimensional convolutions for sliding feature extraction in the temporal dimension, mapping the original sequence to length-invariant local hidden representations, emphasizing short-range textures, instantaneous jumps, and local morphological consistency. Let the convolution kernel size be k and the output channel dimension be C. A non-linear activation function $\phi(\bullet)$ is used to obtain the local representation matrix:

$$H = \phi(Conv1D(X; W_c)) \in R^{T \times C}$$

Where W_c represents the convolutional layer parameters. This represents the input of the original-scale token sequence to the subsequent cross-scale modeling module, while retaining the time-by-time localization capability to support the time-aligned output of anomaly scoring.

3.3 Transformer modeling across time scales

To characterize cross-period and long-range dependencies, a multi-scale sequence is constructed from the local representation H, and tokens with longer receptive fields are obtained by using different downsampling rates. Let the scale set be K, and each scale $k \in K$ correspond to a downsampling rate r_k. Temporal pooling is used to obtain the scale sequence.

$$Z^{(k)} = Pool_{r_k}(H) \in R^{\lceil T/r_k \rceil \times C}$$

Self-attention is applied at each scale to model global dependencies, and the attention is computed as follows:

$$Attn(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d}}\right)V$$

Q, K, V is obtained by a linear mapping, and d is the attention channel dimension. The multi-scale output is then aligned back to the original resolution via temporal upsampling so that it can be fused with the local representation at the same temporal index.

3.4 Fusion and output head

The local representation is fused with the aligned cross-scale context to form a time-by-time joint representation, and anomaly scores and sequence-level risk probabilities are generated separately by a lightweight discriminant head. Let the aligned cross-scale context be $\widetilde{Z} \in R^{T \times C}$, and the fusion uses concatenation and linear mapping:

$$U = LN([H \| Z]W_f) \in R^{T \times D}$$

Where $LN(\bullet)$ is the layer normalization and W_f is the fusion mapping parameter. The time-by-time anomaly score is defined as:

$$s_t = \sigma(w_s^T u_t + b_s),\ t = 1, \ldots, T$$

Where u_t is the t-th row vector of U, and $\sigma(\bullet)$ is the Sigmoid function. Sequence-level risk discrimination obtains the global representation $\tilde{u}$ through temporal aggregation and outputs the class distribution:

$$p = softmax(W_r \bar{u} + b_r),\ \bar{u} = \frac{1}{T}\sum_{t=1}^{T} u_t$$

Where p is the risk category probability vector and W_r, b_r is the discriminant head parameter.

4. Experimental Results and Analysis

4.1 Dataset

This paper uses free historical market data from Stooq as the research dataset, constructing a multivariate time series sample consistent with the task of financial risk anomaly detection and discrimination. Stooq provides historical market data covering multiple categories of assets, including stocks, indices, forex, commodities, and crypto assets, and supports batch or single-asset downloads in CSV format, facilitating the formation of a unified data pipeline and reproducible data slices.

In terms of data organization, representative indices and their related assets are selected to form a cross-market feature set. Based on fields such as daily OHLCV, sequence features such as returns, volatility proxy, and trading activity are derived, resulting in a dimensionally consistent input matrix for anomaly scoring and risk category discrimination. To ensure time alignment and cross-asset consistency, all assets are resampled, and missing data are handled using the same trading calendar. The same notation and field specifications are used for data cleaning and feature construction.

4.2 Experimental setup

The experimental setup revolves around a unified training process for time-series anomaly detection and risk assessment. The input consists of aligned multivariate financial sequence fragments. The model comprises a one-dimensional convolutional local encoder and a cross-timescale Transformer. The output includes time-by-time anomaly scores and sequence-level risk category probabilities. Data preprocessing includes time alignment, missing value handling, and feature standardization. A fixed random seed is used during training to ensure reproducibility, and unified evaluation and comparison are performed under the same inference settings.

The hardware and software environment and key training hyperparameters are summarized in Table 1, including the operating platform, deep learning framework version, optimizer, and learning rate settings, batch size, and sequence length. This configuration balances training stability and computational efficiency, and is consistent with the multi-scale token construction required for cross-timescale modeling.

Table 1: Hardware/software environment and main hyperparameter settings

Category	Item	Setting
Hardware	GPU	NVIDIA RTX 4090 24GB
Hardware	CPU	Intel Xeon, 16 cores
Hardware	Memory	64 GB
Software	OS	Ubuntu 22.04 LTS
Software	Python	3.10
Software	Framework	PyTorch 2.2
Software	CUDA	12.1
Training	Precision	Mixed precision
Training	Optimizer	AdamW
Training	Base learning rate	0.0001
Training	Weight decay	0.01
Training	Batch size	64
Training	Sequence length	256
Training	Feature dimension	32
Training	Conv kernel size	7
Training	Conv channels	128
Training	Transformer layers	4
Training	Attention heads	8
Training	Hidden dimension	256
Training	Dropout	0.1
Training	Gradient clipping	1.0
Training	Epochs	200
Training	Early stopping patience	20
Training	Random seed	42

4.3 Experimental Results and Analysis

To systematically compare the performance of different methods in the task of detecting and identifying anomalies in financial time series risks, representative works with similar research directions to this paper were selected for comparison. Table 2 summarizes the comparison settings of relevant methods under a unified evaluation index system, which is used to comprehensively evaluate them from the dimensions of detection capability, discrimination capability, and prediction reliability.

Table 2. Experimental results compared with other models

Method	Accuracy	Precision	Recall	F1	AUROC	AUPRC	ECE	Brier
Mu et al.[8]	0.841	0.832	0.823	0.824	0.875	0.851	0.041	0.093
Biriukova et al.[9]	0.861	0.854	0.849	0.842	0.887	0.862	0.039	0.089
Kim et al.[10]	0.873	0.862	0.857	0.853	0.888	0.874	0.037	0.088
Wan et al.[11]	0.884	0.871	0.863	0.861	0.909	0.889	0.036	0.086
Li et al.[12]	0.895	0.889	0.872	0.878	0.911	0.898	0.034	0.084
Zeng et al.[13]	0.881	0.875	0.861	0.866	0.902	0.886	0.035	0.085
Berti et al.[14]	0.872	0.863	0.854	0.857	0.897	0.877	0.038	0.087
Ours	0.934	0.921	0.912	0.912	0.965	0.941	0.021	0.062

The comparison in Table 2 shows that the proposed method exhibits a stronger overall discriminative ability, with classification-related indicators surpassing all control methods. This indicates that the combination of local pattern extraction and cross-timescale dependency modeling can more fully characterize the multi-morphological features of risk anomalies. This trend is consistent across multiple indicators, suggesting that the performance improvement does not stem from accidental fluctuations in a single metric, but rather from an overall improvement in feature representation and discrimination boundaries.

From the perspective of detection and ranking capabilities, curve-related indicators show a simultaneous improvement, reflecting that the model's differentiation between abnormal and normal samples is more stable under different threshold settings, and its detection of minority risk signals is more comprehensive. The differences between control methods are relatively limited, mainly reflected in insufficient characterization of complex temporal dependencies and cross-period structures. The proposed method, through cross-scale context fusion, can establish a more consistent discriminative basis between short-term disturbances and medium- to long-term structural changes, thereby reducing confusion.

In terms of reliability, the calibration and error-related indicators are superior, indicating a higher degree of matching between the output probability and the actual risk occurrence tendency, and stronger usability of the risk score. This result means that the model can not only provide more accurate judgments, but also provide more credible confidence information, which is conducive to reducing the decision-making costs caused by excessive alarms and underreporting in risk warning and tiered handling scenarios, and improving the robustness of the overall risk control process.

The learning rate determines the step size of parameter updates and the stability of the optimization trajectory, and is a key factor affecting the convergence pattern and generalization ability of temporal risk discrimination models. For structures that include local convolutional representations and cross-timescale dependency modeling, different learning rates will change the degree of synergy between local pattern extraction and global correlation learning, thus affecting the model's distinction boundary between abnormal and normal samples. To verify the robustness of the training process to learning rate perturbations, it is necessary to systematically scan the learning rate while keeping other settings constant and observe the pattern of performance changes accordingly.

Figure 2 shows a single-peak shape, rising initially and then falling back, indicating a significant bandwidth effect of the learning rate on the model's discriminative ability. At smaller learning rates, parameter updates are limited, the optimization process is more conservative, and the model's capture of key temporal patterns and boundary shaping is relatively insufficient. As the learning rate increases, gradient updates become more thorough, and the synergy between local feature extraction and cross-scale dependency modeling is more easily established, leading to performance improvements.

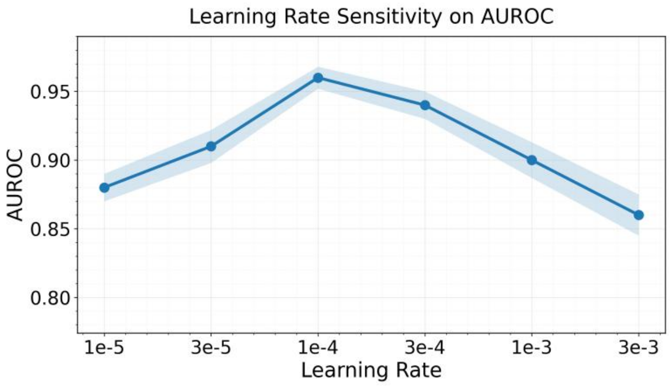

Figure 2. Sensitivity experiment of learning rate to AUROC

When the learning rate continues to increase to a higher range, the curve shows a continuous decline, reflecting that excessively large update steps can disrupt training stability and weaken generalization performance. This phenomenon is usually related to gradient noise amplification, parameter oscillations, and inconsistent convergence paces between different modules. Especially in structures that simultaneously include convolutional local encoding and cross-temporal-scale attention modeling, overly rapid updates may cause global dependency learning to be pulled by local perturbations, resulting in suboptimal representations.

The learning rate range near the peak of the curve exhibits better robustness, and its performance is relatively insensitive to small perturbations, making it suitable as a default setting or a starting point for parameter tuning. Conversely, within the range at both ends of the curve, the performance changes more drastically with the learning rate, indicating that this range is more sensitive to training dynamics and is more susceptible to data noise and initialization differences in practical applications.

The number of convolutional channels determines the representational capacity and fine-grained pattern capture capability of the local encoder, and is a crucial structural hyperparameter affecting the separability of temporal anomalous signals. Too few channels may limit the diverse expression of local features, while too many channels may introduce redundant representations and increase optimization difficulty, thereby altering the efficiency of utilizing context dependencies in the cross-timescale modeling stage. To verify the model's stability under different representational capacities, the number of convolutional channels was scanned in tiers while keeping other settings constant, and the discriminative ability was observed to change with the capacity. The experimental results are shown in Figure 3.

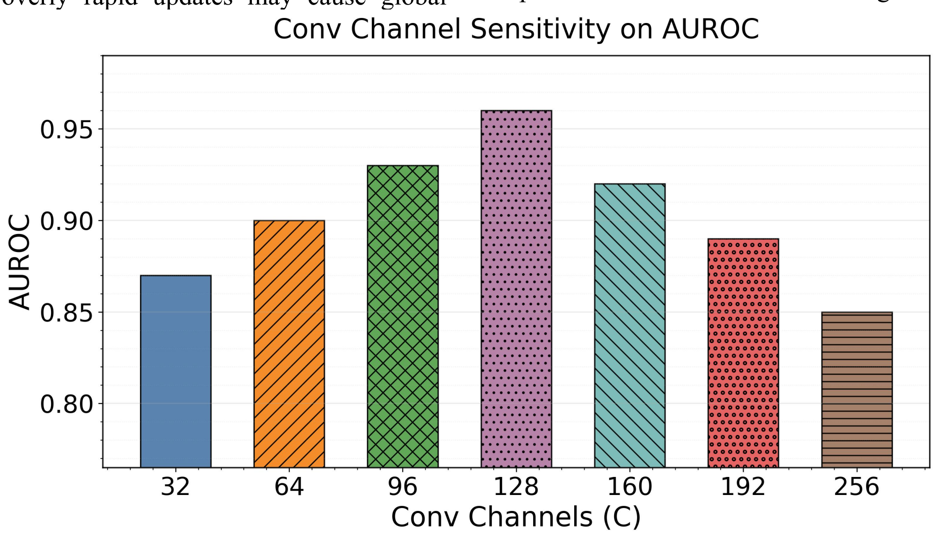

Figure 3. Sensitivity experiment of the convolution channel number to AUROC

Experimental results show that the representation capacity of the local encoder has a clear optimal range for discriminative ability. With a smaller number of channels, the model is limited in its characterization of local fluctuation patterns and short-range mutations, easily leading to insufficient expression of fine-grained cues related to anomalies. As the number of channels increases, the diversity and discriminative power of local features improve, enabling subsequent cross-timescale modeling to obtain higher-quality input representations and thus more effectively form risk boundaries.

Furthermore, performance declines as the number of channels continues to increase, indicating that the increased representation redundancy and optimization difficulty offset the benefits of increased capacity. A larger number of channels not only expands the parameter space but may also amplify the proportion of noisy features in the local encoding stage, making cross-scale attention more susceptible to interference from irrelevant patterns when aggregating context, thereby weakening the focus on key anomaly cues. In addition, the convergence pace between local and global modules may become inconsistent, leading to a decrease in the effectiveness of fused representations.

The channel size near the peak of the curve achieves a better balance between expressive power and training stability, and has a higher tolerance for capacity changes, making it suitable as the default configuration. In contrast, the capacity settings at both ends are more sensitive to performance; the former is limited by insufficient representation, while the latter is more prone to degradation due to redundancy and instability. This pattern suggests that the selection of the number of convolution channels should be based on a balance between separability and robustness, rather than simply pursuing a larger model capacity.

The dimensionality of input features determines the amount of information available to the model and the degree of redundancy, which is a key factor affecting the separability of risk anomalies and the stability of confidence ranking. Too low a dimensionality may lead to insufficient representation of key signals, while too high a dimensionality may introduce noisy features and weaken the focusing ability of modeling across time scales. To characterize the model's adaptability to changes in feature dimensionality, it is necessary to perform a graded scan of the input dimensions and observe the performance distribution under a fixed network structure and training strategy. The experimental results are shown in Figure 4.

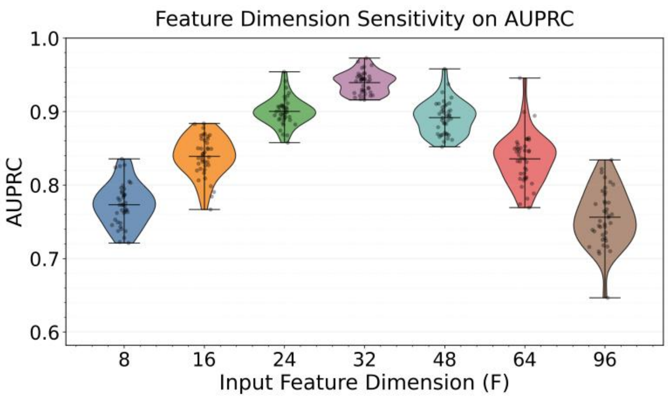

Figure 4. Sensitivity experiment of input feature dimension to AUPRC

The violin distribution exhibits a pattern of initial concentration and upward shift with increasing input dimensionality, followed by a gradual decline and increased dispersion. This suggests that a moderate feature dimension is more conducive to forming stable ranking capabilities and a clearer representation of anomalous signals. In low-dimensional settings, information capacity is limited, resulting in an overall lower distribution and more pronounced fluctuations. As the dimensionality increases, the model can utilize richer multivariate cues, leading to a more concentrated distribution and a robust range near the peak. Further increasing the dimensionality increases the proportion of redundant and noisy features, making the discrimination boundary more susceptible to interference, causing the distribution to shift downward and exhibiting greater variance. This indicates that excessively high dimensionality weakens the focus and stability on key anomalous cues.

5. Conclusion

It constructs a unified modeling framework combining one-dimensional convolution and cross-timescale Transformers, enabling the collaborative characterization of local fluctuation patterns and long-term dependency structures. This framework enhances the capture of short-term mutations and fine-grained patterns with a local encoder, supplements cross-cycle context and stage-specific structural information with cross-scale attention, and achieves alignment and fusion on a unified time axis, ensuring a consistent representational basis for anomaly scores and risk category outputs. The resulting holistic approach better suits the realities of multi-form coexistence, cross-scale coupling, and significant distribution drift in financial risk signals, providing a more robust modeling path for risk warning to move from single-point discrimination to structured identification.

From an application perspective, this research provides a transferable time-series modeling paradigm for scenarios such as transaction monitoring, risk warning, and compliance risk control. For anomaly behavior identification during trading processes, the unified framework can more effectively retain sensitivity to key local patterns in complex noise backgrounds, while leveraging cross-scale context to enhance the identification of persistent risk accumulation and state transitions. For institutional-level risk management and regulatory technology, the model's output anomaly scores and risk assessments can be used for risk stratification, alarm prioritization, and resource scheduling, thereby reducing operational costs from excessive alarms and improving the ability to capture potential systemic risk signals. In the context of increasingly common cross-market linkages and multi-asset allocation, this method's emphasis on consistent cross-asset representation and cross-cycle correlation modeling also helps to more realistically reflect risk transmission paths and multi-scale synergistic effects.

Methodologically, this work demonstrates the complementarity of local convolutional representation and global attention modeling, and enhances its adaptability to non-stationary financial sequences through cross-timescale structural design. The unified framework, which balances anomaly localization and risk attribution discrimination within the same representation space, allows different output tasks to share key temporal structure information, thereby reducing information loss and decision inconsistencies caused by task fragmentation. This approach is not only applicable to traditional market data sequences but also provides a scalable infrastructure for modeling more complex financial signals, including transaction data, order books, and derivative implicit information, which has direct value for building more reliable and scalable intelligent risk control systems.

Future work can be further expanded in three directions. First, to address more complex cross-asset and cross-market relationships, a more granular relationship modeling mechanism can be introduced to explicitly integrate cross-asset linkages and structural changes into cross-scale contextual learning, thereby enhancing the ability to characterize risk transmission and resonance phenomena. Second, to meet the needs for interpretability and auditability in practical deployments, attention attribution, saliency analysis, and rule constraints can be combined to present the connections between key time segments, key features, and risk categories more transparently, thus enhancing the usability and credibility of the model output in regulatory and business decision-making. Third, to address the non-stationary environment of long-term operation, online update and continuous learning mechanisms can be further explored to improve adaptability to distribution drift and institutional changes while ensuring stability, promoting the framework to form a more continuous and reliable risk identification and early warning capability in real financial business.

References

[1] Xu J, Wu H, Wang J, et al. Anomaly transformer: Time series anomaly detection with association discrepancy[J]. arXiv preprint arXiv:2110.02642, 2021.

[2] Tuli S, Casale G, Jennings N R. Tranad: Deep transformer networks for anomaly detection in multivariate time series data[J]. arXiv preprint arXiv:2201.07284, 2022.

[3] Yang Y, Zhang C, Zhou T, et al. Dcdetector: Dual attention contrastive representation learning for time series anomaly detection[C]//Proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining. 2023: 3033-3045.

[4] S. Tuli, G. Casale and N. R. Jennings, "TranAD: Deep Transformer Networks for Anomaly Detection in Multivariate Time Series Data," arXiv preprint arXiv:2201.07284, 2022.

[5] Q. Li, W. Li, C. Zhang, Y. Chen, Y. Wang, Z. Zhang and S. Lu, "Multi-Scale Anomaly Detection for Time Series with Attention-Based Recurrent Autoencoders," Asian Conference on Machine Learning, pp. 674-689, 2023.

[6] A. Deng and B. Hooi, "Graph Neural Network-Based Anomaly Detection in Multivariate Time Series," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, no. 5, pp. 4027-4035, 2021.

[7] Nam Y, Yoon S, Shin Y, et al. Breaking the time-frequency granularity discrepancy in time-series anomaly detection[C]//Proceedings of the ACM Web Conference 2024. 2024: 4204-4215.

[8] K. Cao, "Subgraph-Aware Graph Representation Learning for Collaborative Risk Scoring and Organized Fraud Detection," 2024.

[9] Biriukova K, Bhattacherjee A. Using transformer models for stock market anomaly detection[J]. Journal of Data Science, 2023, 2023(21): 1-8.

[10] Kim J, Kang H, Kang P. Time-series anomaly detection with stacked Transformer representations and 1D convolutional network[J]. Engineering Applications of Artificial Intelligence, 2023, 120: 105964.

[11] L. Shen, Z. Li and J. Kwok, "Timeseries Anomaly Detection Using Temporal Hierarchical One-Class Network," Advances in Neural Information Processing Systems, vol. 33, pp. 13016-13026, 2020.

[12] Z. Li, Y. Zhao, J. Han, Y. Su, R. Jiao, X. Wen and D. Pei, "Multivariate Time Series Anomaly Detection and Interpretation Using Hierarchical Inter-Metric and Temporal Embedding," Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, pp. 3220-3230, 2021.

[13] Zeng Z, Kaur R, Siddagangappa S, et al. Financial time series forecasting using CNN and transformer[J]. arXiv preprint arXiv:2304.04912, 2023.

[14] Y. Zheng, "Modeling Financial Market Dynamics with Temporal and Relational Learning: An LSTM-GNN Approach," 2024.