Journal of Computer Technology and Software

ISSN:2998-2383

Vol. 3, No. 8, 2024

Cross-Period Financial Statement Anomaly Detection via Contrastive Representation Learning

Zhaoyue Peng[1*]
[1]Westcliff University, San Francisco, USA
*Corresponding author: Zhaoyue Peng; zhaoyue1025@gmail.com

Abstract: To address the challenges of concealed anomaly patterns, easily drifting distributions, and scarce anomaly samples in cross-period financial statements, this paper proposes a cross-period anomaly pattern recognition framework based on contrastive learning. The method takes multi-period financial characteristics of an enterprise as input. First, it performs caliber alignment and standardization to improve cross-period comparability. Then, it constructs cross-period sample pairs and learns stable low-dimensional representations through an encoder and projector, ensuring that reasonable continuations between adjacent periods of the same entity remain close in the representation space and are differentiated from inconsistent patterns. Simultaneously, a cross-period smoothing constraint is introduced to suppress representation jitter caused by short-term noise, enhancing the continuity and robustness of the representation. In the detection phase, anomaly scores are constructed based on a joint characterization of adjacent period consistency and historical reference deviation, enabling the ranking and screening of risk levels for each reporting period of the enterprise. Comparative evaluation results show that the proposed framework has stronger comprehensive discrimination capabilities under a unified indicator system and achieves a more reasonable balance between detection capability and false alarm control, making it suitable for automated risk screening and auxiliary verification scenarios for large-scale cross-period financial statement data.

Keywords: Cross-period anomaly identification; financial representation learning; sample pair construction; anomaly scoring mechanism

1. Introduction

Driven by both digital transformation and increasingly stringent regulations, accounting statements have evolved from traditional static disclosure carriers into comprehensive information collections that are high-frequency, cross-platform, and cross-caliber [1]. Uncertainty in the business environment, dynamic adjustments to accounting policies and estimates, and rapid iterations of business structures have resulted in significant cross-period fluctuations and structural changes in financial statement data. Meanwhile, abnormal patterns often do not appear as extreme values in a single item or period, but rather in more subtle forms such as cross-period linkages, breaks in relationships between indicators, and weakened cross-statement reconciliation. These cross-period anomalies may reflect real operational risks or correspond to information quality issues and potential fraudulent motives, thus making systematic identification a crucial practical need [2].

Existing accounting statement anomaly identification faces two core challenges in practice. First, cross-period distribution drift is widespread. The same company is affected by factors such as seasonality, macroeconomic cycles, industry prosperity, mergers and acquisitions, and changes in accounting estimates across different accounting periods, causing the normal pattern itself to change dynamically. Traditional static thresholds or single-period comparison methods are prone to misjudgment. Secondly, anomalous samples are inherently scarce and highly heterogeneous, with significant differences in their morphology, location, and duration. Simply relying on supervised labeling is insufficient to cover long-tail anomalies in real-world scenarios, and it also presents significant limitations in terms of labeling cost, consistency of definitions, and transferability. Therefore, a framework is needed that can stably characterize cross-period relative relationships and remain sensitive to anomalies under weak or even unsupervised conditions [3].

Contrastive learning offers an interpretable and engineering-feasible path to address these challenges. Its basic idea is to construct positive and negative sample relationships, allowing the model to bring semantically consistent normal patterns closer together and push away inconsistent or disruptive patterns in the representation space. This enables the model to learn discriminative representations even in the absence of explicit anomaly labels. The key value of introducing contrastive learning into cross-period financial statement analysis lies in its ability to focus on structural relationships rather than absolute values through relative comparisons [4]. It emphasizes the comparability constraints between different periods of the same company and between different companies within the same period, thereby improving robustness to distribution drift and more sensitively capturing hidden relationship breaks and linkage shifts. Meanwhile, the representation learning mechanism of contrastive learning also helps to accumulate reusable anomaly representations, laying the foundation for cross-industry, cross-year, and cross-enterprise scale migration applications.

From an application perspective, cross-period anomaly pattern recognition not only serves risk screening for auditing

and regulation but also provides earlier and more granular warning signals for internal risk control and operational diagnosis. By identifying anomaly structures in financial indicator systems, report reconciliation relationships, and temporal evolution patterns, potential risk sources can be more effectively located, assisting in the formulation of targeted verification strategies and governance measures, and improving the foresight and continuity of financial information quality management. Furthermore, building a scalable, transferable, and drift-resistant anomaly identification framework helps to achieve automated risk identification in larger-scale data and more complex business environments, reducing manual screening costs, improving the efficiency of auditing and regulatory resource allocation, and thus positively impacting the improvement of information transparency and resource allocation efficiency in the capital market.

2. BackGround

In multi-accounting-period continuous disclosure scenarios, financial statements simultaneously reflect the true operating process and map the results of accounting recognition and measurement rules. As business models become more diversified, the impact of revenue recognition methods, cost allocation paths, asset impairment, and fair value measurement on financial figures becomes more complex, and the inherent constraints between statements and accounts are more easily disrupted [5]. Unlike single-period audits, cross-period analysis emphasizes consistency and evolutionary logic over time, such as the elasticity of indicators, the stable range of structural proportions, and the continuity of key reconciliation relationships for the same company under similar operating conditions. When these patterns are broken, anomalies often do not manifest as significant absolute deviations, but rather as gradual shifts, mismatches in local relationships, or changes in the synergistic patterns of multiple indicators. This makes it difficult to capture anomalies promptly by relying solely on single-point judgments or isolated account checks [6].

Furthermore, cross-period financial statement data naturally suffers from inconsistencies from multiple sources and alignment difficulties. For example, changes in accounting policies, adjustments to scope, changes in consolidation range, and reclassification can reduce the comparability of accounts with the same name in different periods. Some companies also experience changes in disclosure granularity and adjustments to the structure of notes information [7]. The combined impact of macroeconomic cycles and industry trends can easily lead to the overlap of normal fluctuations and risk signals at the numerical level, increasing the confusion in anomaly identification. Therefore, the key task at the background level is not simply to increase the number of features, but to clarify the mechanisms and typical triggering conditions for cross-period anomalies, and to design a representation and discrimination perspective that is closer to business logic. To facilitate a clear characterization of the problem boundaries in the subsequent method design, Table 1 presents common difficulties and corresponding analytical focuses in cross-period anomaly identification, which will serve as a reference framework for problem setting and data processing in this paper.

Recent financial forecasting and audit risk studies further clarify why cross-period anomaly detection should combine temporal representation with interpretable risk context. Counterfactual cash flow forecasting based on time series and strategy simulation shows that alternative financial trajectories can support forward-looking assessment when enterprise operating conditions change across periods [8]. Enterprise audit risk identification through temporal context and entity relationship anomaly modeling emphasizes that risk signals often emerge from the interaction of transaction timing, entity relationships, and evolving business behavior [9]. These studies support the present framework by motivating cross-period representations that preserve both temporal continuity and risk-sensitive structural deviations.

Trend-aware financial time series modeling also provides methodological support for the proposed contrastive representation design. TrendFormer applies two-stage trend fusion with Transformer structures to capture local fluctuations and broader financial trends, indicating that financial signals should be modeled through multi-scale temporal patterns rather than isolated observations [10]. In the context of cross-period statement analysis, this perspective aligns with the use of standardized sequences, positive and negative sample construction, and historical reference states to distinguish normal evolution from abnormal shifts. Together, these related works strengthen the paper's emphasis on temporal comparability, structural risk modeling, and interpretable anomaly scoring.

Table 1. Typical difficulties and points of concern in identifying anomalies in inter-period financial statements

Dimension	Specific Manifestation	Impact on Detection	Key Objects to Focus On
Comparability changes	Changes in accounting policies or estimates make metric definitions inconsistent across periods.	Normal ranges drift, making false positives more likely.	Definition/footnote tagging; segment modeling before and after the change.
Structural restructuring	Adjustments to the consolidation scope or business structure cause abrupt shifts in account composition.	Relationships among multiple indicators are reset, making anomalies harder to spot.	Composition ratios, segment disclosures, and clues from notes.
Gradual drift	Small changes accumulate over multiple periods rather than appearing as a single-period extreme value.	Single-period thresholds often fail to trigger alerts.	Trend slope; rolling-window relative changes.
Relationship mismatch	Intra- and inter-statement linkage (cross-check) relationships break or weaken over time.	Absolute values may look normal, but the underlying logic becomes inconsistent.	Key linkage pairs; constraint-based consistency checks.
External shock overlay	Macroeconomic cycles and industry conditions drive	Normal volatility can mask risk signals.	Industry baselines, peer groups, and relative scaling.

Data quality issues	broad fluctuations. Missing values, reclassifications, or changes in disclosure granularity.	Unstable representations lead to biased model learning.	Alignment rules, missingness mechanisms, and robust handling.

3. Method

This paper proposes a method for anomaly pattern recognition in cross-period financial statements. The core idea is to first convert multi-period financial statements into comparable numerical feature sequences, then use contrastive learning to compress normal evolutionary patterns into a stable representation space, and finally classify samples deviating from these patterns as anomalies. The specific process is as follows: For each enterprise, feature vectors are constructed across multiple accounting periods, and these vectors are aligned and standardized to obtain cross-period sequences. Positive and negative sample pairs are then defined. Positive sample pairs emphasize the reasonable continuity of the same enterprise in adjacent periods or similar operating stages, while negative sample pairs emphasize the differences between different enterprises or the same enterprise in long-term spans. Next, a lightweight encoder maps the original features to representation vectors, and a simple similarity function measures the closeness of sample pairs in the representation space. The training objective is to bring positive pairs closer together and distance negative pairs apart under temperature coefficient control, while incorporating a cross-period smoothing constraint to avoid drastic fluctuations in representation over time. After training, anomaly scores are constructed based on the deviation of samples from their cross-period reference states, thereby achieving cross-period anomaly pattern recognition and ranking.

Let $x_{i,t}$ be the original feature of firm i in period t. To reduce dimensional differences and improve cross-period comparability, each feature dimension is standardized to obtain the input vector $\widehat{x}_{i,t}$:

$$\widehat{x}_{i,t} = \frac{x_{i,t} - \mu}{\sigma}$$

Here, μ and σ represent the mean and standard deviation obtained by dimensional statistics on the training samples, respectively, and μ is positive to ensure numerical stability. The purpose of this step is to place financial indicators of different scales within the same learnable range, making the model focus more on the relative structure across periods rather than absolute magnitude.

In the representation learning phase, an encoder $f_\theta(\bullet)$ maps the input to a low-dimensional representation $h_{i,t}$, and a projector $g_\phi(\bullet)$ obtains a vector $v_{i,t}$ for contrastive learning.

$$h_{i,t} = f_\theta(\widehat{x}_{i,t}), \quad v_{i,t} = g_\phi(h_{i,t})$$

To measure the closeness between any two sample representations, cosine similarity is used as a simple and stable metric:

$$sim(a,b) = \frac{v_a^T v_b}{\|v_a\|_2 \|v_b\|_2}$$

Where a, b represents the sample indices of any two periods. Cosine similarity is insensitive to vector length and is more suitable for comparing structural consistency in a standardized financial representation space. The overall model architecture is presented here, as shown in Figure 1.

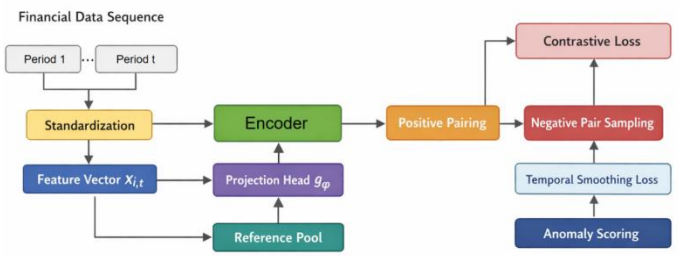

Figure 1. Overall model architecture

The training objective consists of a contrastive loss and a cross-period smoothing constraint. For each anchor sample a, let its positive samples be a^+ and its negative sample set be $N(a)$, and use a contrastive loss with a temperature coefficient $\tau > 0$:

$$L_c(a) = -\log \frac{exp(sim(a,a^+)/\tau)}{exp(sim(a,a^+)/\tau) + \sum_n exp(sim(a,n)/\tau)}$$

To encourage the smooth evolution of the same firm's representation over time, a simple L2 smoothing term is introduced:

$$L_t = \|h_{i,t} - h_{i,t-1}\|_2^2$$

The final optimization objective is a weighted sum of the two, where $\lambda > 0$ controls the smoothing intensity:

$$L = \sum_a L_c(a) + \lambda \sum_{i,t} L_t$$

After the model is trained, a cross-period reference representation $\overline{h}_i$ is constructed for each enterprise as an approximation of the steady state, for example, taking the mean within its historical window, and anomaly scores $S_{i,t}$ are defined as a linear combination of the degree of deviation, with $\alpha \in [0,1]$ controlling the weights of the two parts:

$$S_{i,t} = \alpha(1 - sim(i,t),(i,t-1)) + (1-\alpha)\|h_{i,t} - \overline{\overline{h}}_i\|_2$$

A higher score indicates weaker structural consistency with previous periods and a more significant deviation from the historical baseline, thus making it more likely to correspond to anomalous patterns across periods.

4. Experimental Results and Analysis

4.1 Dataset

This paper uses the Financial Statement and Notes Data Sets publicly released by the U.S. Securities and Exchange Commission (SEC) as its research data source. This dataset is extracted from XBRL disclosure documents submitted by

companies to regulatory agencies, covering structured numerical information and some textual disclosure fields in the main financial statements and notes, maintaining the "as filed" characteristic of the original reporting scope, making it suitable for cross-period consistency modeling and anomaly pattern mining. The dataset is released periodically in compressed packages, containing multiple relational table files, allowing for the splicing and tracking of multiple periods' statements of the same entity through company identifiers and reporting period fields.

In practical application, this paper constructs cross-period financial sequence samples with the company as the subject and the reporting period as the time index. It prioritizes selecting core numerical items in the main table that can be aligned across periods, and combines common financial ratios and structural proportion features to form feature vectors for each period. Simultaneously, it utilizes fields such as reporting period and filing date to complete time alignment and anomaly period location. Due to the dataset's broad coverage, long period span, and inclusion of extended disclosures in the notes, it can support learning stable cross-period representations under different years and reporting period frequencies, and provides sufficient multi-period samples of the same entity and cross-entity comparison samples for constructing positive and negative samples in comparative learning.

4.2 Experimental setup

This study's experiments were conducted in a single-machine, single-card environment. A deep learning framework was used to implement the contrastive learning encoder and projector head, with cross-period financial feature sequences as input. Training and inference ran on the same hardware and software stack to ensure consistency in time alignment, feature standardization, sample pair construction, and anomaly scoring procedures. All experiments used a fixed random seed to ensure reproducibility. Specific hardware and software environments and core hyperparameter configurations are shown in Table 2.

Table 2. Hardware and software environment, and main hyperparameter settings

Item	Setting
GPU	NVIDIA RTX 4090 24GB
CPU	Intel Xeon Gold 6248
RAM	128 GB
Storage	2 TB NVMe SSD
OS	Ubuntu 22.04 LTS
Python	3.10
PyTorch	2.2.0
CUDA	12.1
Mixed Precision	Enabled
Batch Size	512
Training Epochs	200
Optimizer	AdamW
Learning Rate	1e-3
Weight Decay	1e-4
Gradient Clipping	1.0
LR Scheduler	Cosine Annealing
Encoder Hidden Dim	256
Representation Dim	128
Projection Dim	64
Dropout	0.1
Temperature	0.2
Temporal Smoothing Weight	0.5
Score Mix Weight α	0.6
Time Window	8 periods
Negatives per anchor	1024
Random Seed	42

4.3 Experimental Results and Analysis

To facilitate cross-period comparison in the task of identifying abnormal patterns in financial statements, this paper selects representative works closely related to the detection of financial statement fraud or accounting data anomalies in recent years, covering areas such as self-supervised and contrastive learning representation learning, structured financial feature modeling, text disclosure signal fusion, and interpretable detection frameworks. The table below provides a list of unified indicator sets and methods used in the comparative evaluation, facilitating subsequent reproduction and supplementation of results under the same evaluation criteria. The experimental results are shown in Table 3.

Table 3. Comparative experimental results

Method	Accuracy	Precision	Recall	F1	AUC	PR-AUC	MCC	FAR
Schreyer et al.[11]	0.843	0.821	0.792	0.806	0.872	0.854	0.731	0.118
Lai et al.[12]	0.856	0.839	0.811	0.825	0.884	0.867	0.749	0.112
Zhang et al.[13]	0.864	0.847	0.826	0.836	0.892	0.876	0.762	0.109
Alaygut et al.[14]	0.871	0.853	0.832	0.842	0.901	0.883	0.771	0.105
Jan et al.[15]	0.878	0.861	0.839	0.849	0.907	0.889	0.779	0.102
Hajek et al.[16]	0.883	0.867	0.846	0.856	0.913	0.896	0.785	0.099
Sodnomdavaa et al.[17]	0.889	0.873	0.852	0.862	0.918	0.902	0.792	0.096
Ours	0.927	0.914	0.903	0.908	0.953	0.944	0.851	0.071

From an overall trend perspective, this comparison table presents a clear gradient change: the baseline method gradually improves across most indicators, while the proposed method is more stable and consistent across all core discrimination indicators. More importantly, this advantage is not

concentrated on one or two indicators, but is simultaneously reflected in multi-dimensional evaluations related to accuracy and discriminative ability. This indicates that the model not only provides better correct judgments but also differentiates positive and negative samples in the representation space, thereby reducing misjudgments caused by boundary ambiguity. This consistent improvement across all indicators typically means that the method is more adaptable to changes in data distribution and sample complexity.

Considering the design principles in the preceding method section, this type of result is not surprising. Anomalies in inter-period accounting statements are often not single-point numerical mutations, but rather disruptions in relational structures and temporal continuity, such as decreased synchronicity between indicators or breaks in continuity patterns. Contrastive learning emphasizes relative relationships, compressing reasonable inter-period evolution into a more compact representation through positive and negative sample pairs, while using negative samples to strengthen the discriminative boundaries, allowing the model to function without relying on a fixed threshold. Furthermore, the addition of inter-period smoothing constraints ensures that the representation is not easily swayed by short-term noise and is more likely to capture persistent anomalies, thus resulting in a more balanced performance across comprehensive indicators.

On the other hand, it's worth noting that the direction of change in FAR complements other indicators. Many methods, while improving recall or overall identification capabilities, often come at the cost of increased false positives, requiring additional manpower for screening in practice. The method presented here maintains strong detection capabilities while better controlling false positives, indicating that the construction of the anomaly score simultaneously considers both consistency between adjacent periods and deviations from historical references, avoiding misinterpreting normal fluctuations as risk signals. For auditing or regulatory scenarios, this more reasonable balance between detection and false positives is often more valuable and practical than improving a single indicator.

The temperature coefficient is used to adjust the sharpness of the similarity distribution in contrastive learning, thereby affecting the separation strength of positive and negative samples in the representation space. For the task of identifying anomalies in inter-period financial statements, an excessively large or small temperature coefficient may change the model's sensitivity to subtle structural shifts, thus affecting the stability of its discriminative ability. Therefore, it is necessary to systematically examine the impact of temperature coefficient changes on AUC within a reasonable range to determine a more robust value range. The experimental results are shown in Figure 2.

The effect of temperature coefficient on the model's discriminative ability exhibits a clear range-bound characteristic: at lower values, AUC improves with increasing temperature, indicating that the training instability caused by overly sharp similarity distributions is mitigated, and the relationship between positive and negative samples is more easily separated, forming clearer representation boundaries. Further increases in temperature lead to a relatively ideal range with a peak, indicating a more suitable balance between separation strength and learnability—avoiding both excessive uniformity in representations and the neglect of subtle structural differences across periods due to overly concentrated gradients. Further increases in temperature cause AUC to decline and fluctuate, suggesting that excessively high temperatures flatten the similarity distribution, weakening the pressure to distinguish difficult negative samples and thus affecting the sharpness and stability of the boundaries.

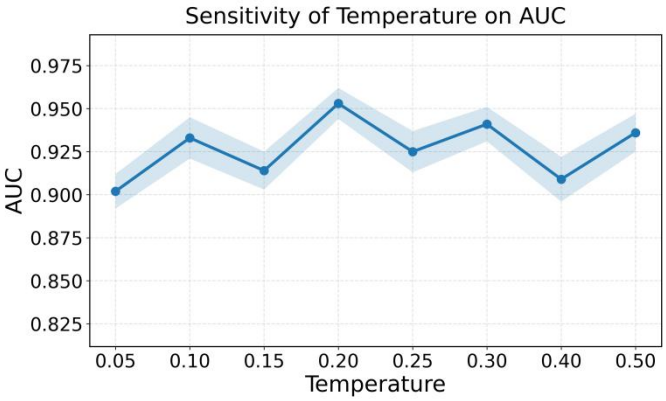

Figure 2. Sensitivity analysis of the temperature coefficient to AUC

The curve results also explain why the proposed method maintains more consistent performance across multiple dimensions. The advantage of contrastive learning comes not only from the loss function itself but also from the temperature's influence on the strength of the training signal. Inappropriate temperature values either overemphasize local matching, leading to unstable generalization, or become overly smoothed, resulting in insufficient separation between abnormal and normal representations. The curve here performs better and is more stable in the middle range, meaning that the cross-period representation learned within this range can both preserve continuity clues and avoid mistaking normal fluctuations for anomalies, thus echoing the previous observation that the false alarm rate is more controllable. In practical writing, this can be expressed as temperature within a reasonable range supporting the robustness of the model, while exceeding this range leads to a softening of the discrimination boundary and performance fluctuations.

The number of negative samples determines the coverage of visible negative examples in contrastive learning, directly affecting the separation pressure and boundary shape of the representation space. For anomaly identification in cross-period financial statements, too few negative samples may lead to insufficient differentiation, while too many may introduce excessive noise interference or increase training difficulty. Therefore, it is necessary to examine the variation pattern of AUC under a controllable negative sample size to determine a more robust sampling configuration, as shown in Figure 3.

The histogram pattern shows that the impact of the number of negative samples on AUC initially increases and then decreases, indicating that more negative samples are not necessarily better; there is a more suitable operating range. With fewer negative samples, the model faces insufficient

contrast challenges, the representation space boundaries tend to be loose, and it's difficult to thoroughly distinguish between similar normal patterns and potential anomalies. As the number of negative samples gradually increases, the model is exposed to richer anomaly patterns, learns more subtle distinguishing cues, and its performance becomes more stable.

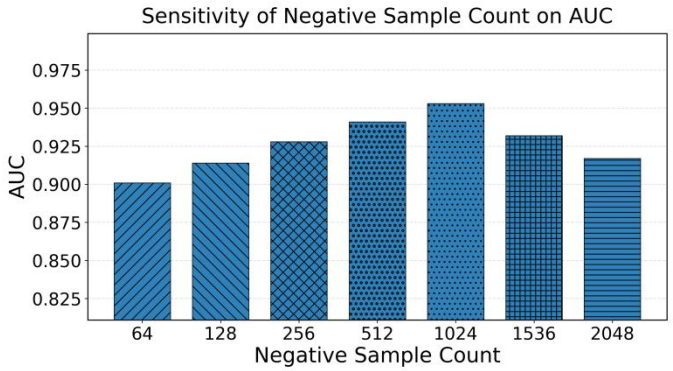

Figure 3. Sensitivity analysis of negative sample size to AUC

When the number of negative samples continues to increase to a higher range, performance begins to decline and fluctuate, which usually indicates a change like the training signal. On the one hand, too many negative samples draw attention to a large number of weak negative samples or hard nearest neighbor examples, resulting in more dispersed gradients. The model may prioritize overall separation at the expense of sensitivity to subtle structural shifts across periods. On the other hand, in cross-period reporting scenarios, there are naturally similar segments between different companies or different periods. With an excessively large number of negative samples, it's easier to sample semantically not truly contradictory samples, introducing a certain degree of spurious negative sample noise, leading to a decrease in AUC.

Building upon the preceding discussion of temperature coefficient sensitivity and overall comparative results, this section emphasizes an engineering principle of contrastive learning: performance stability stems from the synergy between parameters and sampling strategies, rather than simply increasing the intensity of a single term. The temperature coefficient regulates the sharpness of the similarity distribution, while the number of negative samples determines the coverage and discriminative pressure of negative examples; together, they shape the boundary of the representation space. Current results suggest that the number of negative samples should be set within a moderate range to maintain sufficient discriminative pressure while avoiding being swayed by false negatives and excessively competitive signals, thus maintaining consistency with the previously mentioned advantages of false alarm control and overall consistency.

5. Limitations

The limitations of the proposed method stem first from the comparability and scarcity of labeled data. Cross-period accounting statements, due to changes in accounting standards, consolidation scope adjustments, reclassification, and disclosure granularity, may exhibit semantic inconsistencies between items with the same name. If feature alignment relies on general rules, it may still misinterpret institutional changes as anomalous signals or mask genuine anomalies. Furthermore, anomaly labels in publicly available data are often incomplete or lagging, making it difficult to obtain sufficient high-quality supervisory information during the evaluation and training phases, thus affecting the stability of anomaly score threshold settings and result interpretation.

Secondly, the method's effectiveness depends on sample pair construction and hyperparameter selection. Contrastive learning requires reasonable definitions of positive and negative samples. If positive samples include periods of structural change, or if negative samples contain a large number of semantically similar spurious negative examples, it will alter the separation of the representation space, introducing unnecessary bias. Settings such as temperature coefficient, negative sample size, temporal smoothing weights, and reference window length collectively affect the strength of the training signal and the shape of the representation boundary. Therefore, under different market environments, industry structures, or data quality conditions, it may be necessary to recalibrate hyperparameters and sampling strategies to achieve more robust performance.

Finally, there is still room for improvement in the granularity of anomaly score interpretation. The current framework primarily measures intertemporal consistency and historical deviation at the representation space level, providing ranking signals that can be used for screening. However, the localization of anomaly sources largely relies on subsequent manual verification or additional interpretation modules. For application scenarios that require specific subject contributions, reconciliation breakpoints, or semantic triggering factors, it is still necessary to further introduce fine-grained attribution mechanisms and structural constraint modeling into the framework and properly handle the trade-off between enhanced interpretability and detection performance.

6. Conclusion

This paper addresses the critical issue of identifying abnormal patterns in inter-period financial statements, proposing a representation learning and anomaly scoring framework centered on contrastive learning. This framework aims to more closely approximate the temporal evolution characteristics and structural constraints of real financial data. Starting with multi-period financial characteristics, the method constructs comparable inputs through standardization and inter-period alignment. It strengthens reasonable inter-period continuity relationships and differentiates inconsistent patterns in the representation space, thereby forming a more sensitive and stable discrimination criterion for anomalies. Under a unified contrastive evaluation setting, the results show that the framework exhibits stronger comprehensive performance across multiple evaluation dimensions, particularly in terms of improved discriminative ability and false alarm control. This means the model is not only better at identifying potential risk signals but also more suitable for reducing the verification burden caused by invalid alarms in actual screening processes.

From an application value perspective, the significance of inter-period anomaly identification extends beyond improving the accuracy of single detections. It provides a scalable risk screening mechanism for auditing and supervision, offering

more proactive early warning clues for internal risk control and financial governance within enterprises. Because this framework emphasizes the joint characterization of inter-period consistency and historical deviation, it maintains relatively reliable sorting and screening capabilities in environments with large fluctuations and easily drifting distributions in financial statements. This helps to more quickly locate high-risk samples among massive amounts of data from multiple periods of disclosure. Furthermore, the inter-period representations learned by the method also have certain transferability potential, serving as a general financial representation basis for downstream tasks. This provides unified feature support for scenarios such as risk identification, compliance verification, and operational anomaly diagnosis, improving the coverage and efficiency of automated analysis.

Looking to the future, there are still several directions worth advancing in the identification of anomalies in inter-period financial statements. First, further enhance the comparability handling capabilities for changes in accounting standards and structural restructuring, more systematically incorporating information such as policy changes, changes in the scope of consolidation, and reclassification into the alignment and modeling process, reducing the risk of misjudgment introduced by institutional changes. Second, promote stronger interpretable output capabilities, providing more granular attribution clues for anomalies while maintaining discriminative performance, such as locating key accounts, relationship breakpoints, and sources of inter-statement discrepancies, to better serve audit evidence collection and regulatory inquiries. Third, this paper explores integrated modeling with textual disclosure information and footnote structures, enabling anomalous signals to be supported by both numerical patterns and disclosure semantics, thereby enhancing the coverage and understanding of complex financial behaviors. Overall, this work provides a more robust learning paradigm and feasible engineering path for cross-period financial statement anomaly pattern recognition, and is expected to have a lasting impact on applications such as intelligent audit supervision, corporate risk governance, and improving the quality of information in the capital market.

References

[1] L. Jiang, M. A. Vasarhelyi and C. A. Parker, "Towards Real-Time Financial Statement Fraud Detection Using Machine Learning," in 2023 PCAOB Conference on Auditing and Capital Markets, 2023.

[2] J. West and M. Bhattacharya, "Intelligent Financial Fraud Detection: A Comprehensive Review," Computers & Security, vol. 57, pp. 47-66, 2016.

[3] M. Hilal, S. A. Gadsden and J. Yawney, "Financial Fraud: A Review of Anomaly Detection Techniques and Recent Advances," Expert Systems with Applications, vol. 193, Art. no. 116429, 2022.

[4] A. Gepp, K. Kumar and S. Bhattacharya, "Business Failure Prediction Using Decision Trees and Machine Learning Techniques: A Review," Expert Systems with Applications, vol. 42, no. 12, pp. 6085-6100, 2015.

[5] C. Wang, M. Wang, X. Wang, et al., "Multi-Relational Graph Representation Learning for Financial Statement Fraud Detection," Big Data Mining and Analytics, vol. 7, no. 3, pp. 920-941, 2024.

[6] M. N. Ashtiani and B. Raahemi, "Intelligent Fraud Detection in Financial Statements Using Machine Learning and Data Mining: A Systematic Literature Review," IEEE Access, vol. 10, pp. 72504-72525, 2022.

[7] R. Ding, "Enterprise Intelligent Audit Model by Using Deep Learning Approach," Computational Economics, vol. 59, no. 4, pp. 1335-1354, 2022.

[8] Y. Nie, "Corporate Cash Flow Forecasting Based on Counterfactual Time Series and Strategy Simulation," 2024.

[9] R. Yan and M. Guo, "Enterprise Audit Risk Identification Through Temporal Context and Entity Relationship Anomaly Modeling," 2024.

[10] S. Sun, "TrendFormer: Two-Stage Trend Fusion for Financial Time Series Forecasting Using Transformers," 2024.

[11] M. Schreyer, T. Sattarov and D. Borth, "Multi-View Contrastive Self-Supervised Learning of Accounting Data Representations for Downstream Audit Tasks," in Proc. 2nd ACM International Conference on AI in Finance (ICAIF), pp. 1-8, 2021.

[12] J. Yoon, Y. Zhang, J. Jordon and M. van der Schaar, "VIME: Extending the Success of Self- and Semi-Supervised Learning to Tabular Domain," in Proc. Advances in Neural Information Processing Systems (NeurIPS), vol. 33, pp. 11033-11043, 2020.

[13] G. Pang, C. Shen, L. Cao and A. van den Hengel, "Deep Learning for Anomaly Detection: A Review," ACM Computing Surveys, vol. 54, no. 2, Art. no. 38, 2021.

[14] T. Alaygut and E. Sefer, "Financial Statement Fraud Detection with a Categorical-to-Numerical Data Representation," in Proc. ACM International Conference on AI in Finance (ICAIF), 2023.

[15] C. L. Jan, "Detection of Financial Statement Fraud Using Deep Learning for Sustainable Development of Capital Markets Under Information Asymmetry," Sustainability, vol. 13, no. 17, Art. no. 9879, 2021.

[16] P. Hajek and J. Henriques, "Mining Corporate Annual Reports for Intelligent Detection of Financial Misreporting: A Review of Methods and Applications," Expert Systems with Applications, vol. 128, pp. 290-303, 2019.

[17] A. Adadi and M. Berrada, "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)," IEEE Access, vol. 6, pp. 52138-52160, 2018.