Transactions on Computational and Scientific Methods | Vol. 5, No. 11, 2025
ISSN: 2998-8780
https://pspress.org/index.php/tcsm
Pinnacle Science Press

Anomaly Detection in E-Commerce Multi-Table ETL Processes through Mamba-LSTM Collaborative Modeling

Ye Zhang[1], Ethan Cheng[2]
[1]Cornell Tech, New York, USA
[2]Arizona State University, Tempe, USA
*Corresponding author: Ye Zhang; yz2778@cornell.edu

Abstract: To address the challenges of complex anomaly types, tight inter-table relationships, significant temporal dependencies, and the difficulty of traditional methods simultaneously considering global context and local dynamic changes in e-commerce multi-table ETL processes, this paper proposes a data anomaly detection method based on Mamba and LSTM collaborative modeling. First, this method constructs a unified sequence representation for multi-source business data such as orders, payments, logistics, products, and users, mapping heterogeneous field information, relational information, and process state information to a shared feature space to enhance the model's ability to characterize complex business semantics. Building upon this, a Mamba branch is introduced to model long-range dependencies and cross-stage contexts in the ETL chain, improving the ability to identify hidden anomaly propagation patterns. Simultaneously, an LSTM branch is used to capture local state transitions and short-term dynamic change features, enhancing the perception of fine-grained anomaly perturbations. Subsequently, an adaptive fusion mechanism dynamically coordinates the global and local representations to form a unified feature representation for anomaly detection, which is then combined with a classification head to complete anomaly sample identification. This method effectively adapts to the characteristics of multi-source heterogeneity, coupled relationships, and continuous process evolution in e-commerce ETL data, enabling efficient characterization and stable discrimination of abnormal patterns in complex business processes. The results show that the proposed method demonstrates good overall performance in anomaly identification, feature representation sufficiency, and model stability, providing effective support for data quality governance and anomaly monitoring in e-commerce scenarios.

Keywords: E-commerce data governance; multi-table ETL; anomaly detection; time series representation learning

1. Introduction

With the continuous growth of e-commerce platform transaction volume, business data such as orders, payments, logistics, products, and user behavior exhibit significant multi-source heterogeneity [1, 2]. These data are interconnected through primary keys, foreign keys, timestamps, and business status, forming complex and close relationships. Against this backdrop, data processing for e-commerce multi-table scenarios is no longer limited to single-table cleaning and aggregation, but relies more heavily on extraction, transformation, and loading processes to unify, structurally align, and semantically map scattered data [3]. The ETL process, as a crucial foundation supporting e-commerce data warehouse construction, business analysis, intelligent decision-making, and risk control, directly impacts the accuracy and reliability of downstream data applications. Anomalies in multi-table joins, field mapping, status synchronization, or time-series connections can lead to a chain of problems such as distorted metrics, broken links, and misjudgments in business processes. Therefore,

research on data anomaly detection in e-commerce multi-table ETL processes has significant practical needs and theoretical value.

Compared to traditional single-source data processing scenarios, e-commerce multi-table ETL data anomalies exhibit stronger concealment, coupling, and dynamism [4]. On the one hand, anomalies often don't manifest as a single field going out of bounds or missing data, but rather as cross-table semantic deviations such as inconsistencies between order and payment tables, conflicts between logistics tracking and transaction status, misalignments between product information and customer feedback, and incomplete user behavior chains. On the other hand, e-commerce businesses are characterized by high concurrency, strong timeliness, and continuous evolution. Platform promotions, inventory changes, payment delays, and refund repatriations constantly alter data distribution, causing anomaly patterns to exhibit significant temporal fluctuations and contextual dependencies. Traditional methods relying on static rules, manual verification, or shallow statistics often struggle to effectively characterize the deep-seated relationships between multiple tables and are ill-suited to the continuously changing anomaly patterns under complex business processes. Therefore, there is an urgent need to introduce novel methods capable of simultaneously modeling temporal dependencies and long-range associations [5].

In recent years, deep learning methods focused on sequence modeling and complex dependency capture have provided new research avenues for ETL anomaly detection. Among these, the cyclic modeling mechanism has a natural advantage in handling local time dependencies and continuous state evolution, and can better reflect the dynamic processes in e-commerce business processes such as order status progression, payment behavior response, and logistics event updates [6]. Meanwhile, the novel state-space modeling mechanism shows strong potential in preserving long-sequence information, expressing global dependencies, and improving computational efficiency, providing important support for handling long-distance associations across stages, links, and table entries in e-commerce multi-table ETL. Organically integrating these two modeling approaches is expected to simultaneously address both fine-grained local dynamic changes and global long-range dependency expressions, thereby improving the ability to identify complex anomaly patterns. Based on this, research on anomaly detection in e-commerce multi-table ETL data using the Mamba and LSTM fusion framework not only aligns with the development trend of intelligent data governance methods but also provides a new technical path for complex business data quality management.

From an application perspective, constructing anomaly detection methods for e-commerce multi-table ETL scenarios is of great significance for improving platform data governance, ensuring the credibility of business decisions, and enhancing the stability of business systems. This type of research not only helps to identify problems such as data extraction errors, transformation biases, missing loads, inconsistencies between tables, and process interruptions promptly, reducing the risks of anomalous data propagating to downstream tasks, but also provides a more reliable data foundation for real-time monitoring, precise operation, financial verification, supply chain collaboration, and user service optimization for e-commerce platforms [7]. From an academic perspective, unifying the characteristics of multi-table ETL processes, cross-table relationships, and temporal evolution into an anomaly detection modeling framework helps to promote data quality management research from static detection of single tables to dynamic perception of multiple tables, and provides valuable theoretical support for intelligent ETL, automated data operation and maintenance, and anomaly governance of complex data systems.

2. BackGround

E-commerce platform data systems typically consist of multiple business tables, including order tables, payment tables, product tables, user tables, logistics tables, and review tables. Different tables record specific states within the transaction process and are linked through order numbers, user numbers, product numbers, time fields, and business status fields. This multi-table organization can comprehensively reflect the entire e-commerce process from browsing, ordering, payment, shipping, to receipt and review, but it also presents data processing challenges such as complex structure, dense relationships, and frequent updates. In data warehouse

construction and business analysis, ETL undertakes key tasks such as raw data access, field transformation, data cleaning, table joins, and result loading. Its execution quality directly determines the consistency and usability of the analysis topic. Due to the high-frequency interaction and continuous change characteristics of e-commerce, ETL processes not only need to handle large-scale time-series data but also need to ensure semantic and state-level consistency between different business tables, which places higher demands on subsequent anomaly identification and data quality control.

In multi-table ETL scenarios, anomalies often do not occur in isolation but rather propagate and accumulate along multiple stages of the business chain. For example, source table field drift can trigger mapping errors, which in turn can lead to state mismatches, failed joins, or missing loads, causing anomalies to exhibit a clear chain reaction of effects. Meanwhile, e-commerce business data also suffers from strong time dependencies, context sensitivity, and ambiguous anomaly boundaries. The same field value may have different meanings at different business stages, and the same combination of states may correspond to either normal or abnormal results under different time windows. Therefore, relying solely on rule matching or local statistics is insufficient to fully reflect the potential relationships between data from multiple tables. To more accurately describe the complex changes in the ETL process, it is necessary to understand the formation mechanism of e-commerce data anomalies from three levels: temporal evolution, cross-table joins, and contextual semantics. This will lay the foundation for subsequently building an anomaly detection model with deep representation capabilities.

3. Methodology

To characterize the heterogeneity of multi-table ETL records in e-commerce systems, the proposed method first converts the original relational stream into a unified event sequence that preserves transaction semantics, temporal order, and cross-table dependency patterns. Each ETL instance is represented as a sequence of fused event tokens derived from order, payment, logistics, commodity, and user-related tables, so that the model can jointly perceive attribute values, linkage states, and process evolution under a common feature space. This paper presents the overall model architecture, as shown in Figure 1.

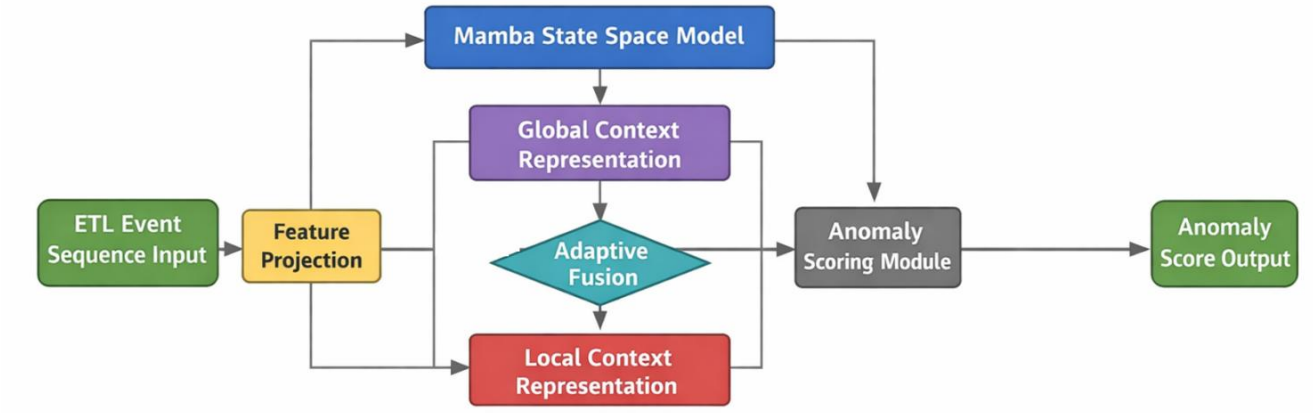


Figure 1. Overall model architecture

Rather than treating abnormality as an isolated fluctuation in a single table, the modeling strategy emphasizes the coupled deviation accumulated across extraction, transformation, and loading stages, because data inconsistency in practical ETL pipelines often emerges through delayed propagation, key mismatch, status conflict, or incomplete synchronization. Given an input sample of length T, the sequence is denoted by:

$$\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_T]$$

where $\boldsymbol{x}_t \in \mathbb{R}^d$ integrates numerical attributes, categorical embeddings, temporal descriptors, and relational indicators extracted at the t th ETL step. To enhance structural awareness before temporal modeling, a learnable projection is introduced to map heterogeneous raw features into a compact latent representation through:

$$\boldsymbol{h}_t^0 = \phi(\boldsymbol{W}_e \boldsymbol{x}_t + \boldsymbol{b}_e)$$

in which $\boldsymbol{W}_e$ and $\boldsymbol{b}_e$ denote trainable parameters and $\phi(\cdot)$ serves as a nonlinear transformation for improving representation continuity across different tables. Such a design is essential because ETL anomalies rarely follow identical numerical patterns, whereas they frequently share latent relational signatures after semantic alignment. A positional temporal encoding is then injected to retain order sensitivity and process continuity, which is written as:

$$\boldsymbol{z}_t = \boldsymbol{h}_t^0 + \boldsymbol{p}_t$$

with $\boldsymbol{p}_t$encoding step-wise temporal context, thereby allowing the subsequent network to distinguish whether a suspicious state occurs at an early extraction stage or during later loading and reconciliation stages.

Beyond simple sequential aggregation, long-range dependency modeling is necessary because many ETL abnormalities are only visible when distant operations are jointly inspected. For this reason, a Mamba-based state space branch is employed to capture global dependency propagation with linear-time scanning behavior, and its hidden dynamics are formulated as:

$$\boldsymbol{s}_t = \boldsymbol{A}_t \boldsymbol{s}_{t-1} + \boldsymbol{B}_t \boldsymbol{z}_t$$

where the latent state $\mathbf{s}_t$ accumulates historical process information, while input-dependent matrices $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$ adaptively control memory retention and feature injection according to the current ETL context. Global contextual output is further generated by:

$$\boldsymbol{m}_t = \boldsymbol{C}_t \boldsymbol{s}_t$$

so that the representation can reflect whether a current tuple-level observation is compatible with the broader cross-table execution trajectory. This branch is particularly meaningful for multi-table ETL analysis because a loading anomaly may originate from an earlier extraction distortion, and a status inconsistency may only become detectable after several intermediate transformations have altered its local appearance.

Complementing the global branch, an LSTM-based local evolution module is introduced to preserve short-term transition regularity and fine-grained state continuity, since adjacent ETL actions often contain strong causal cues regarding duplication, omission, or abrupt mutation. The local memory update is expressed as:

$$\boldsymbol{c}_t = \boldsymbol{f}_t \odot \boldsymbol{c}_{t-1} + \boldsymbol{i}_t \odot \tilde{\boldsymbol{c}}_t$$

where the forget gate $\boldsymbol{f}_t$ suppresses irrelevant historical residue, the input gate $\boldsymbol{i}_t$ determines how much new evidence should be absorbed, and $\tilde{\boldsymbol{c}}_t$ denotes the candidate memory generated from the current observation and preceding hidden state. Local sequential output is then obtained by:

$$\boldsymbol{l}_t = \boldsymbol{o}_t \odot tanh(\boldsymbol{c}_t)$$

with $\boldsymbol{o}_t$ regulating the exposed anomaly-sensitive pattern at time step t. Because ETL data quality problems often manifest as subtle local disturbances before evolving into system-level inconsistency, the LSTM branch provides a complementary perspective that is more responsive to nearby state transitions, incremental drift, and transient operational irregularities.

After the two branches produce global and local descriptors, an adaptive fusion mechanism is adopted to reconcile broad dependency awareness with short-range transition sensitivity, thereby preventing the model from overemphasizing either delayed contextual evidence or immediate local variation. Fusion weights are learned dynamically through:

$$\boldsymbol{\alpha}_t = \sigma(\boldsymbol{W}_a[\boldsymbol{m}_t; \boldsymbol{l}_t] + \boldsymbol{b}_a)$$

where $\sigma(\cdot)$ is the sigmoid activation, $[\cdot\,;\cdot]$ denotes vector concatenation, and $\boldsymbol{\alpha}_t$ measures the relative contribution of the two branches under the current ETL state. The final anomaly-aware representation is defined as:

$$\boldsymbol{u}_t = \boldsymbol{\alpha}_t \odot \boldsymbol{m}_t + (1 - \boldsymbol{\alpha}_t) \odot \boldsymbol{l}_t$$

so that abnormal patterns caused by long-chain semantic inconsistency and those induced by local operation disorder can be modeled within a unified space. Based on the aggregated sequence representation, the anomaly score is finally produced by a prediction head and optimized with a supervised objective tailored to binary ETL abnormality identification. This design improves methodological interpretability because the decision process is no longer tied to a single temporal scale, but instead emerges from coordinated reasoning over latent process memory, short-term update behavior, and cross-table semantic coherence.

4. Experimental Results and Analysis

4.1 Dataset

The Olist Brazilian E-Commerce Public Dataset is a publicly available, real-world multi-table dataset of e-commerce transactions. Originating from the Brazilian e-commerce platform Olist, it covers nearly 100,000 orders and their associated business information from approximately 2016 to 2018, exhibiting strong multi-table relationships and complete business process information. This dataset typically includes several core relational tables such as order tables, order product tables, order payment tables, order review tables, customer tables, seller tables, product tables, and geolocation tables. It can describe a relatively complete transaction chain from order placement, payment, and shipment to receipt and review, making it highly suitable for researching multi-table ETL, data integration, relationship consistency checks, and anomaly detection in e-commerce scenarios. Because the tables are explicitly linked through fields such as order number, customer number, product category, seller number, and postal code prefix, this dataset not only supports order-level time-series analysis but also supports research on cross-table field matching, status synchronization checks, missing relationship identification, duplicate record cleaning, and business rule validation. Therefore, compared to single-table transaction data, it better reflects the complexity and application value of real-world e-commerce data warehouses and ETL processes.

Since the Olist dataset does not provide explicit labels for ETL anomaly detection, the anomaly labels used in this study are derived during the data preprocessing and multi-table integration stage. Specifically, after joining the order, payment, product, customer, and other related tables, each integrated record is examined according to relational integrity constraints and business consistency rules. Records with complete associations, valid temporal order, and consistent numerical and status information are assigned the normal label, whereas records exhibiting broken table associations, missing linked fields, duplicated records, inconsistent status transitions, abnormal payment aggregation, temporal conflicts, or incomplete business chains are assigned the anomalous label.

4.2 Experimental Results and Analysis

To further illustrate the effectiveness of our proposed method in anomaly detection for multi-table ETL data in e-commerce, we compared it with representative deep learning methods in time series anomaly detection and multivariate anomaly recognition. These methods cover different technical approaches, including recurrent neural networks, autoencoder reconstruction, generative modeling, and Transformer modeling, comprehensively reflecting the current development of related research. Based on this, we present a comparative evaluation under a unified metric, tailored to our research scenario, to illustrate the differences in detection accuracy and overall discriminative ability among various methods. The experimental results are shown in Table 1.

Table 1. Experimental results compared with other models

Method	Precision	Recall	Acc	AUC
Malhotra et al. [8]	88.42	85.76	84.91	85.33
Munir et al. [9]	89.37	86.45	85.72	86.08
Geiger et al. [10]	90.14	87.62	86.83	87.22
Audibert et al. [11]	91.08	88.41	87.95	88.18
Su et al. [12]	91.56	88.97	88.34	88.65
Tuli et al. [13]	92.47	90.12	89.66	89.89
Xu et al. [14]	92.83	90.58	89.94	90.26
Ours	94.31	92.47	93.38	92.17

The proposed algorithm demonstrates strong comprehensive recognition capabilities across multiple evaluation metrics, indicating that it can more fully uncover the potential differences between anomalous and normal samples and maintain high stability and consistency in the anomaly detection process. These results show that the constructed model exhibits good adaptability in feature representation, temporal relationship modeling, and anomaly pattern capture, effectively improving the anomaly detection quality in e-commerce multi-table ETL scenarios. For data mismatches, state conflicts, and hidden anomalies in complex business chains, the method demonstrates strong sensitivity and robustness, further reflecting the rationality of the model design.

Furthermore, while accurately identifying anomalous samples, the proposed method also ensures complete coverage of anomaly information, reflecting that the model does not rely solely on local features for judgment but can complete anomaly representation learning within a richer contextual semantics. This indicates that the introduced key mechanism can effectively alleviate the shortcomings of traditional methods in complex sequence modeling and deep association capture, enabling the model to maintain superior performance even when facing multi-source heterogeneity, strong temporal dependencies, and inter-table coupling relationships in e-commerce ETL data. In summary, the algorithm presented in this paper has high application value and provides reliable methodological support for subsequent data quality monitoring and anomaly early warning for real business processes.

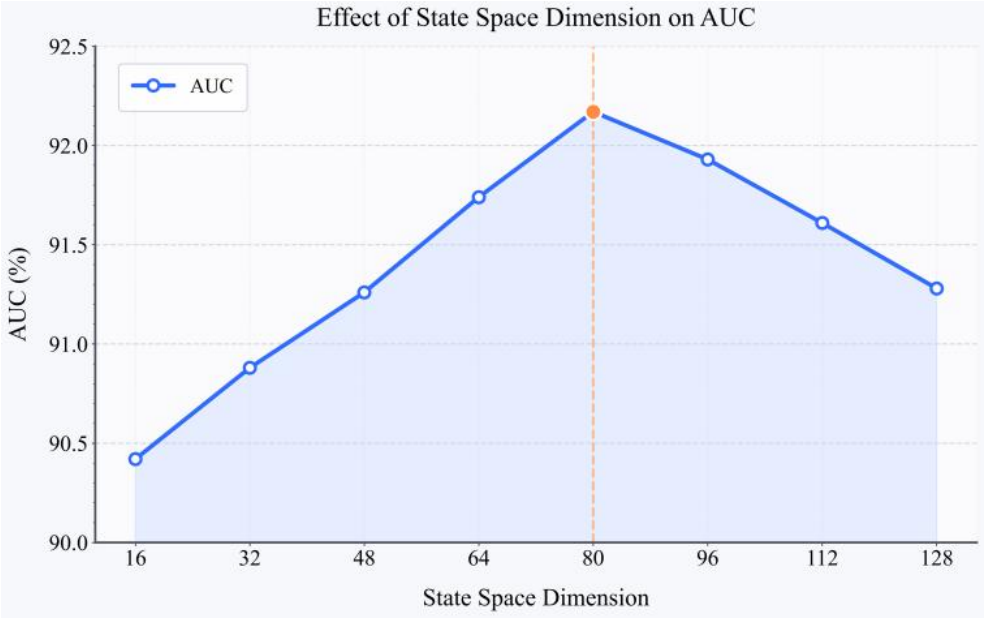


Figure 2. The impact of state-space dimensionality sensitivity on AUC

As shown in Figure 2, the state space dimension setting directly affects the model's representation quality of anomaly features in multi-table ETL for e-commerce. The proposed algorithm maintains good discriminative ability under this parameter configuration, indicating strong adaptability in global dependency modeling and anomaly information extraction. Since multi-table ETL data inherently possesses complex temporal correlations and cross-table semantic coupling characteristics, key clues in anomaly patterns can only be more completely preserved when the latent states have sufficient information-carrying capacity. Therefore, this result further demonstrates the rationality of the proposed method's structural design.

The algorithm effectively coordinates global contextual memory and local dynamic changes during state space representation, enabling anomaly identification to move beyond shallow fluctuations at a single level and achieve discrimination from deeper business relationships. This shows that the constructed Mamba and LSTM collaborative mechanism can provide a more stable representation foundation for complex ETL links, and also demonstrates the high application potential of the proposed method in addressing issues such as multi-source heterogeneity, continuous evolution of e-commerce processes, and blurred anomaly boundaries, providing reliable support for subsequent data quality monitoring and anomaly early warning tasks.

5. Conclusion

This paper focuses on data anomaly detection in e-commerce multi-table ETL scenarios. Addressing the limitations of traditional methods, such as insufficient cross-table join modeling, limited long-range dependency characterization, and inadequate expression of complex business semantics, this paper proposes an anomaly detection method based on collaborative modeling using Mamba and LSTM. Starting from the real characteristics of e-commerce business processes, this method integrates the temporal relationships and semantic connections formed by multi-source data (orders, payments, logistics, products, and users) during the ETL process into a unified modeling framework. Through the synergistic effect of global context awareness and local dynamic evolution characterization, the model's ability to identify complex anomaly patterns is enhanced. The results show that data quality issues in multi-table ETL links cannot be judged solely based on single-table statistical characteristics or local rules; instead, a comprehensive analysis from multiple levels, including business process continuity, inter-table consistency, and the rationality of state evolution, is necessary. The detection framework built around this approach provides a more systematic technical path for e-commerce data governance and further enriches the research content in the field of ETL anomaly detection.

From a methodological perspective, the significance of this paper lies not only in proposing an anomaly detection model suitable for multi-table ETL scenarios in e-commerce, but also in advancing anomaly detection research from single-source static analysis to multi-source temporal collaborative modeling. E-commerce business data is characterized by high update frequency, strong structural heterogeneity, complex inter-table coupling relationships, and concealed anomaly manifestations. This makes anomaly detection inherently possess multiple attributes, including sequence understanding, semantic alignment, and relation inference. The proposed method considers both long-distance business dependencies and nearest-neighbor state changes during the modeling process, enabling anomaly identification to be built on a more complete process semantic foundation, thereby improving the model's ability to perceive potential risks in complex business chains. This research approach has strong inspirational significance for data warehouse construction, data platform governance, and intelligent ETL operation and maintenance, and also illustrates that data anomaly detection for real business processes needs to pay more attention to the collaborative expression between temporal logic and relational structure.

From an application value perspective, this research has strong practical significance for improving the quality of e-commerce platform data assets, ensuring the accuracy of business analysis, and enhancing the operational stability of business systems. In real-world business environments, if an ETL process encounters errors such as field mapping errors, state synchronization conflicts, broken correlation links, or loading deviations, anomalous data can propagate along the analysis chain, impacting multiple critical aspects, including business decisions, inventory management, financial verification, user profiling, and risk identification. The method

proposed in this paper provides more intelligent and refined support for data quality monitoring in complex e-commerce scenarios, shifting anomaly detection from post-incident investigation to process awareness and proactive identification, thereby reducing the cascading impact of data issues on business systems. Furthermore, this research can provide transferable technical references for related fields such as retail supply chain collaboration, intelligent warehouse scheduling, precision marketing analysis, and digital operations management, thus possessing broad application potential.

Future research can be further deepened in several directions. First, by combining more granular business rules and domain knowledge, order fulfillment logic, payment status constraints, logistics timeliness relationships, and user behavior context can be further integrated into the modeling process to enhance the business interpretability and scenario adaptability of anomaly detection. Secondly, online detection mechanisms for real-time ETL streams can be explored, enabling models to have stronger dynamic update and continuous early warning capabilities in continuously arriving data environments, thereby meeting the higher timeliness requirements of high-concurrency e-commerce platforms. Thirdly, it can be further extended to data governance scenarios involving cross-platform, multi-regional, and multi-business system collaboration, improving the model's generalization performance and deployment value in heterogeneous environments. Finally, as enterprise data infrastructure continues to evolve towards intelligence and autonomy, ETL anomaly detection methods for complex business processes will not only undertake data cleaning and quality assurance functions, but will also gradually become an important component supporting the stable operation, intelligent decision-making, and risk prevention of digital enterprises.

References

[1] J. Song, K. Kim, J. Oh et al., "MemTO: Memory-Guided Transformer for Multivariate Time Series Anomaly Detection," Advances in Neural Information Processing Systems, vol. 36, pp. 57947-57963, 2023.

[2] Y. Yang, C. Zhang, T. Zhou et al., "DCdetector: Dual Attention Contrastive Representation Learning for Time Series Anomaly Detection," Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 3033-3045, 2023.

[3] Z. Liu, X. Huang, J. Zhang et al., "Multivariate Time-Series Anomaly Detection Based on Enhancing Graph Attention Networks With Topological Analysis," Proceedings of the 33rd ACM International Conference on Information and Knowledge Management, pp. 1555-1564, 2024.

[4] H. Sun, Y. Huang, L. Han et al., "HCL-MTSAD: Hierarchical Contrastive Consistency Learning for Accurate Detection of Industrial Multivariate Time Series Anomalies," arXiv preprint arXiv:2404.08224, 2024.

[5] C. S. Lee, "Long-Range Dependency Modeling and Decision Point Summarization for Large Language Models in Dialogue and Meeting Scenarios," 2024.

[6] Z. Liu, R. Meng, S. Y. Huang and Z. Huang, "Cost-Sensitive Mamba Sequence Modeling for Fault Detection in Cloud-Native Microservice Systems," 2025.

[7] J. Kou, W. Wang and Y. Xu, "Collaborative Decision Optimization for Timely Order Fulfillment and Service Quality Enhancement in E-Commerce Supply Chains," 2025.

[8] P. Malhotra, A. Ramakrishnan, G. Anand et al., "LSTM-Based Encoder-Decoder for Multi-Sensor Anomaly Detection," arXiv preprint arXiv:1607.00148, 2016.

[9] M. Munir, S. A. Siddiqui, A. Dengel et al., "DeepAnT: A Deep Learning Approach for Unsupervised Anomaly Detection in Time Series," IEEE Access, vol. 7, pp. 1991-2005, 2018.

[10] A. Geiger, D. Liu, S. Alnegheimish et al., "TadGAN: Time Series Anomaly Detection Using Generative Adversarial Networks," Proceedings of the 2020 IEEE International Conference on Big Data, pp. 33-43, 2020.

[11] J. Audibert, P. Michiardi, F. Guyard et al., "USAD: Unsupervised Anomaly Detection on Multivariate Time Series," Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 3395-3404, 2020.

[12] Y. Su, Y. Zhao, C. Niu et al., "Robust Anomaly Detection for Multivariate Time Series Through Stochastic Recurrent Neural Network," Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 2828-2837, 2019.

[13] 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.

[14] J. Xu, H. Wu, J. Wang et al., "Anomaly Transformer: Time Series Anomaly Detection With Association Discrepancy," arXiv preprint arXiv:2110.02642, 2021.