International Journal of Emerging Trends in Engineering and Technology
ISSN: 3107-8699 | Volume 2 Issue 1, 28-32, January-March 2026
DOI: https://doi.org/10.64137/31078699/IJETET-V2I1P105

Received: 15/12/2025 Revised: 22/12/2025 Accepted: 13/01/2026 Published: 18/01/2026

Original Article

Self-Supervised Anomaly Detection in Logistics Quality Data Streams

G. RAJAGANESAN
Department of Commerce, Thanthai Periyar Government Arts and Science College, Tiruchirappalli, Tamil Nadu, India.

ABSTRACT: The increasing digitalization of logistics operations has resulted in continuous streams of quality-related data generated from sensors, automation systems, and information platforms. Traditional anomaly detection approaches often rely on labeled datasets or static rules, which are impractical in dynamic logistics environments where quality deviations are rare, evolving, and context-dependent. This study proposes a self-supervised anomaly detection framework for monitoring logistics quality data streams in real time. The proposed approach learns normal operational patterns directly from unlabeled data by leveraging self-supervised learning techniques, enabling the detection of handling, transportation, storage, and sensor-related anomalies without prior fault annotations. The framework incorporates adaptive anomaly scoring and streaming data processing to address concept drift and operational variability. Experimental results demonstrate that the proposed model outperforms conventional unsupervised and supervised baselines in terms of detection accuracy, robustness, and early warning capability. The findings highlight the potential of self-supervised learning as an effective tool for proactive quality assurance in modern logistics and supply chain systems.

KEYWORDS: *Self-Supervised Learning, Anomaly Detection, Logistics Quality Management, Supply Chain Analytics, Streaming Data, Artificial Intelligence, Quality Engineering.*

1. INTRODUCTION

Most traditional quality monitoring methods in logistics are based on a rule-based approach or supervised machine learning models. They operate using fixed thresholds and expert knowledge in rule-based systems which cannot adapt to changing operational contexts. Although more flexible, supervised learning methods require training on large amounts of labeled anomaly data. Labeled datasets are difficult, expensive and time-consuming to acquire because quality failures in logistics environments are rare and heterogeneous. In addition, supervised models typically do have difficulty with the detection of anomaly patterns that have never been observed prior to the deployment of the predictive system, which renders their application unfit for logistical systems in real-world dynamic environments.

Self-supervised learning has recently emerged as a promising alternative for anomaly detection in environments characterized by abundant unlabeled data and scarce fault annotations. By designing learning tasks that allow models to learn normal behavior directly from raw data, self-supervised methods eliminate the dependency on labeled anomalies. This capability is particularly well suited to logistics quality data streams, which are continuous, high-dimensional, and subject to frequent operational changes. The motivation for this study lies in leveraging self-supervised learning to build adaptive, scalable, and robust anomaly detection systems capable of operating in real time within logistics environments.

Despite the potential of self-supervised approaches, their application to logistics quality monitoring remains underexplored. The core research problem addressed in this study is how to design an effective self-supervised anomaly detection framework that can learn normal logistics quality patterns, detect deviations in real time, and remain robust to noise and concept drift. The primary objective of this research is to develop and validate a self-supervised learning-based system for anomaly detection in logistics quality data streams, focusing on both methodological rigor and practical applicability. The key contributions of this study include the development of a real-time anomaly detection architecture, the integration of self-supervised learning with streaming data processing, and an empirical evaluation demonstrating improved detection performance over conventional approaches.

2. LITERATURE REVIEW

Logistics and supply chain quality monitoring has traditionally been based on inspection, statistical process control, and key performance indicator tracking. These methods focus on analyzing events after the fact or performing periodic audits, and are often too slow to detect quality deviations when they occur. Due to the rise of sensors, enterprise systems, and automation technologies it has been possible to start researching data-driven monitoring techniques that allow for continuous supervision over logistics processes.

Anomaly detection methods have been broadly studied in different areas such as manufacturing, cybersecurity, and finance. Traditional approaches to anomaly detection include statistical modelling, distance-based method, clustering algorithms and rule-based heuristic. These approaches are easy to implement but often assume data distributions remain stationary and struggle with the high-dimensionality, noise, and non-linearity endemic to logistics systems. Such methods have limited scalability and adaptability as logistics operations become increasingly complex.

Machine Learning (ML) based anomaly detection algorithms can be primarily classified into: supervised, unsupervised, and self-supervised methods. Supervised methods are based on labeled datasets, which set apart normal and abnormal behavior types; These methods provide high accuracy when there is enough labeled data. For these unlabeled methods, like clustering and density estimation, anomaly points within the data are detected using either distance (distance-based method) or probability or statistics values of point patterns with respect to all other objects in the dataset as standard measures.Anecdotal research indicates that most unsupervised methods lack strong robustness and observed anomalous objects tend to exist near ambiguous structure objects. Self-supervised learning is in the middle, where a surrogate learning tasks are formed on which the model learns meaningful representations of normal behavior on unlabeled data, self-supervised learning also makes it an ideal unsupervised anomaly detection method for complex systems.

In addition, more challenges are imposed by logistics data streams, which are not well addressed in the existing anomaly detection literature. The challenges include high velocity of data, heterogeneous data sources, temporal dependencies, missing values and the concept drift caused due to changes in operational conditions. Existing research typically works with static datasets and offline analysis, which makes them less suitable for monitoring logistics in real-time. There exists a concrete research gap in crafting an anomaly detection framework that fuses self-supervised learning and real-time stream processing, while taking into account logistics-specific quality requirements. We situate this study to undertake the developing demand in emergent real-time quality data streams by proposing a new framework for customizable, self-supervised anomaly detection specifically for logistics quality data.

3. PROBLEM DEFINITION AND SYSTEM ARCHITECTURE

A logistics quality data stream is a flow of time-stamped data from complementary logistic operations for aspects comprised by product specification, product handling, transportation procedures, storage methods and system performance. These data streams are often multivariate and come among different sources, for example sensors, automation equipment design and Enterprise Information Systems. Logistics quality data streams such as these must immediatley analysed continuously in order to detect any later-than-imminent quality deviations, unlike standard batch datasets.

There are many delivery logistics quality anomalies. Errors due to improper loading, unloading or shocks during movement can cause handling-related anomalies. These include transportation anomalies, such as route deviations, delays and environmental violations (for instance temperature or humidity excursions). Storage based Anomalies Storage anomalies can be changes in warehouse conditions, stock stays idle for a long time or storage configurations. As faulty devices, calibration errors or data transmission issues may lead to sensor-level anomalies that mask or mimic a real quality problem. And in order to enable accurate diagnosis and corrective actions, an anomaly detection system should be able to distinguish between aforementioned types of anomalies.

Logistics quality monitoring have highly heterogeneous sources of data. IoT devices deliver realtime environment and condition-monitoring data, automation systems produce operational metrics, while enterprise platforms like ERP & Warehouse Management Systems capture transactional and process-level details. There are various technical and analytical complexities in the integration of all these data sources into a single monitoring framework.

In order to mitigate these challenges, the proposed system architecture supports real-time anomaly detection via the integration of logistics data streams. The architecture has modules that gather the streaming inputs (data ingestion), preprocess layers to remove the noise and extract potentially significant features, self-supervised learning engine for modeling normal behavior, and an

anomaly detection module which scores in real-time and raises alerts 언제. Such a modular design guarantees scalability, adaptability and interoperability of existing logistics information systems.

4. METHODOLOGY

4.1. DATA PREPROCESSING AND FEATURE ENGINEERING

Data preprocessing is an essential prerequisite for efficiently detecting anomalies in logistics quality data streams. Due to sensor errors, failure of communication or integration of various systems, raw logistics data is often noisy and has missing values. The process also involves data normalization techniques to enable variables of varying scales and ranges interact properly during the learning. Smoothing and filtering as noise handling methods are used to limit the impact of transient fluctuations that do not indicate true quality issues.

Feature engineering extracts meaningful representations from raw data that capture temporal and contextual aspects of logistics operations. Temporal feature extraction refers to modeling the time-dependent patterns, trends, and seasonality of logistics processes. Contextual features include operational conditions, such as location, transport mode or process stage that provide additional context to classify normal variability versus anomalous behavior. Collectively, these components create a diverse input embedding for self-supervised learning architectures.

4.2. SELF-SUPERVISED LEARNING FRAMEWORK

The self-supervised learning framework is designed to learn normal logistics quality patterns without requiring labeled anomaly data. This is achieved by formulating pretext tasks that encourage the model to capture the underlying structure of the data. Pretext tasks may involve reconstructing input sequences, predicting masked or future values, or distinguishing between temporally consistent and inconsistent data segments. These tasks force the model to learn informative representations that reflect normal operational behavior.

The model architecture is selected based on the characteristics of logistics quality data streams. Autoencoder-based architectures are commonly used for their ability to compress and reconstruct high-dimensional data, making reconstruction errors a natural indicator of anomalies. Contrastive learning approaches focus on learning discriminative representations by comparing positive and negative sample pairs. Transformer-based models are particularly effective in capturing long-range temporal dependencies and complex interactions among variables. The chosen architecture balances detection accuracy, computational efficiency, and real-time applicability.

4.3. ANOMALY SCORING AND DETECTION STRATEGY

The output of the self-supervised model is fed into an anomaly scoring algorithm, which scores it as a number (or vector) quantified by how far away from normal behaviour it is. The scores measured by reconstruction-based method is the input itself and its model reconstruction; if the error is large, it indicates that this sample might be anomaly. Similaritybased scores quantify the proximity of newly observed instances/observations to learned representationsnormal. These scores offer a way to measure abnormality in a continuous manner, as opposed to with binary classifications.

Streaming environments, on the other hand, use adaptive thresholding mechanisms to segregate between regular variations and actual anomalies. Static thresholds are frequently unproductive because operating circumstances vary. Adaptive thresholds are where so that the threshold can adapt its value based on recent data distributions and thus allowing for higher sensitivity while reducing false alarms. This method provides a strong way to detect anomalies under dynamic logistics environments.

4.4. REAL-TIME STREAM PROCESSING MECHANISM

Logistics operations require real time stream processing to detect anomalies in a timely way. Sliding Window methods represent a way to break down streaming, continuous data into smaller time units to analyze each separately. These are the windows that make it possible for the model to work with very recent data while keeping their contextual awareness. Strategies of online learning enable the model to update the parameters incrementally over time after introduction or arrival of new data whilst also enhancing its adaptability and long-term performance.

The stream processing mechanism is primarily centered around concept drift handling. Concept drift refers to the phenomenon where the statistical properties of data change because logistics processes evolve, seasonal effects change or operational configurations are altered. Furthermore, the proposed framework presents an integrated solutions for drift detection and model adaptation to evolve representations that are still relevant over time. This approach allows the system to adapt well to new patterns, and consequently enables effective anomaly detection in continuously changing logistics environments.

TABLE 1 Components and functional architecture of a self-supervised learning framework for real-time logistics quality anomaly detection

Component	Description
Logistics Quality Data Streams	Continuous, multivariate, time-stamped data generated from logistics operations, including handling, transportation, storage, and system-level quality indicators collected in real time.
Data Sources	Internet of Things (IoT) sensors, automated material handling systems, transportation monitoring devices, Enterprise Resource Planning (ERP) systems, and Warehouse Management Systems (WMS).
Data Preprocessing	Cleaning raw data by handling missing values, noise, and inconsistencies, followed by normalization to ensure comparability across heterogeneous quality variables.
Feature Engineering	Extraction of temporal features capturing trends and patterns over time, along with contextual features reflecting operational conditions such as location, transport mode, and process stage.
Self-Supervised Learning Strategy	Learning normal operational behavior using unlabeled data through pretext tasks such as sequence reconstruction, masked value prediction, or temporal consistency modeling.
Model Architecture	Deep learning models including autoencoders, contrastive learning networks, or transformer-based architectures designed to capture complex temporal dependencies in logistics data streams.
Anomaly Scoring Mechanism	Quantification of deviations from learned normal patterns using reconstruction error or representation similarity scores to identify potential quality anomalies.
Thresholding Approach	Adaptive thresholding techniques that dynamically adjust decision boundaries based on recent data distributions to reduce false alarms in non-stationary environments.
Stream Processing Method	Real-time analysis using sliding window mechanisms and online learning to process continuous data streams efficiently.
Concept Drift Handling	Detection and adaptation mechanisms that update the model in response to changes in logistics processes, demand patterns, or operational configurations.

5. EXPERIMENTAL DESIGN

The experiment design is oriented on a strict test of the proposed self-supervised quality monitoring solution in conditions that a realistic logistics anomaly classification has. This study could use a simulation or real-world logistics dataset (a set of data collected during the transportation, warehousing and handling operations) as long as it reproduces quality logistic conditions. In real-world, multivariate time-series data exist as a practical case of where datasets such as temperature; humidity; vibration; transit duration or city dwell time (i.e., the amount of time an inventory remains in same place); equipment status—with all collecting from IoT sensors or enterprise systems. In the absence of sufficient task-representative real-world data, and to ensure reproducibility and rigorous experimental control, simulated datasets are generated using extensible stochastic models informed by domain knowledge that captures normal operations along with controlled anomaly injections.

We use multiple evaluation metrics to measure the anomaly detection performance. For evaluating the classification capability of the model for detecting quality anomalies against false alarms, precision, recall and F1-score are used. Detection Delay: A major metric to assess how effectively the model can detect anomalies after their appearance, which is required by early warning systems within logistics quality management. These metrics are a combination of how accurate, robust, and timely the system is.

Baseline models: Once you have the baseline model selection to perform a meaningful comparison, base models are your first computer vision models which results contain meaningful performance to compare with self-supervised approach proposed in the paper. Such baselines can be for instance conventional statistical inferences, classical unsupervised anomaly detection algorithms and supervised learning models that are derived directly or indirectly from previously labelled data. By measuring performance against one or more set of baseline categories, we confirm that any observed improvement is due to the self-supervised learning strategy itself and did not simply arise by increasing the model complexity. Case 1 scenarios seek to encapsulate the nature of many real-world logistics conditions, including operating in an equilibrium state with stable business functions for extended periods (e.g., business-as-usual), gradual changes to processes over time boundaries, and sudden disruptive shocks that precipitate more severe process interventions than would be needed otherwise. Using temporal cross-validation combined with rolling-window evaluation ensures that performance is generalized across both time and operational contexts.

6. RESULTS AND DISCUSSION

The results show that the presented self-supervised anomaly detection model shows a solid performance on all evaluation metrics. It was able to correctly predict quality anomalies with high precision and recall (wrongly flagging operational variability as quality deviation or vice versa) which proves that the model is capable of distinguishing genuine quality deviations from routine

operational variability. The short detection delay we observe in all experiments indicates that this framework gives an early warning, and hence is able to prevent future downstream quality failures.

In experiments on real-world datasets the proposed method outperforms baseline methods significantly, especially when the anomaly pattern is unseen or changes over time. Models based on traditional statistics and unsupervised techniques can lead to excessive false alarms, while supervised models are subject to failure when anomalies deviate from training labels. The self-supervised model learning strong features of normal behavior generalizes better under non-stationary nature which gives it an upper hand on overall performance.

Such anomalies are investigated further to understand their operational impact. The model is able to accurately detect shocks related to handling, transportation delays, environmental threshold violations, and sensor malfunctions. These interpretations correspond to actual logistics quality problems, and thus show that the detected anomalies are not random statistical outliers, but relevant warnings for operational risk in real-world practice. This talk highlights how these insights can facilitate proactive decision making and enable continuous quality improvement in logistics systems.

7. MANAGERIAL AND INDUSTRIAL IMPLICATIONS

From a managerial perspective, the proposed framework provides significant advantages to quality engineers and supply chain managers through automating processes for continuous quality monitoring without the dependence on large labeled datasets. The system eases the manual inspection workloads thereby improving visibility of logistics operations empowering managers to make corrective interventions before small deviations turn into large-scale quality escapes.

It integrates with logistics infrastructure and existing systems such as IoT platforms, warehouse management systems, and enterprise resource planning systems. This allows it to be implemented in practice without making fundamental changes. The anomaly detection system is integrated directly into existing digital processes, which improves operational transparency and decision support for the organization.

Self-supervised anomaly detection not only provides theoretical benefits, but also concrete cost impact, efficiency and risk-reduction benefits. When quality problems are caught early, the product waste, recall costs and service disruptions are minimized. Reducing downtime and optimizing the use of resources increases operational efficiency. The framework also facilitates risk mitigation by increasing resilience to unplanned disruptions, which is especially important in the challenging and volatile environments of many supply chains.

8. LIMITATIONS AND FUTURE RESEARCH DIRECTIONS

The proposed study possesses certain limitations that need to be kept in mind despite its merits. Data availability continues to be one of the major challenges related with access to high quality, reality-grounded datasets, as they are often limited due to restrictions on data confidentiality and security. Simulated data is a partial answer, but some real operational environments are much too complex to be captured in simulation.

In particular, when the framework is implemented on large-scale and heterogeneous logistics networks, scalability and generalization can be problematic. Model performance can vary drastically with differences in data quality, system configurations and operational practices. Future Work Should Focus On Scalable Architectures And Federated Learning Approaches For Enhance Generalization Cross-Organization.

From this framework, one important direction for future work is to extend the workshop scope to multi-tier supply networks. By modeling interactions and dependencies across suppliers, manufacturers, and distributors, it is easy to identify cascading quality risks and defect propagation. Other areas that can improve trust, interpretability in industrial setup and adoption are explainable AI principles and embedding domain knowledge learning principles.

9. CONCLUSION

This research proposes a novel self-supervised learning-based framework to anomalously search logistics quality data streams, which overcome the restrictions that the traditional rule-based and supervised methods have. The proposed framework enables a readable, non-parametrically accurate, fast and adaptive detection of quality anomalies in dynamic logistics environments by allowing model to learn normal operational patterns directly from unlabeled data.

It establishes real-time, scalable and robust anomaly detection architecture for streaming data and therefore contributes to Quality engineering research area and AI-driven logistic systems. The experimental results illustrate better performance than the traditional methods, indicating the practical significance of self-supervised learning in logistics quality management. To sum up, the study emphasizes how state-of-the art AI methods can change quality assurance and make modern supply chain systems more resilient.

REFERENCES

[1] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly Detection: A Survey," *ACM Computing Surveys*, vol. 41, no. 3, pp. 1–58, Jul. 2009, doi: https://doi.org/10.1145/1541880.1541882.

[2] L. Ruff *et al.*, "A Unifying Review of Deep and Shallow Anomaly Detection," *arxiv.org*, Sep. 2020, doi: https://doi.org/10.1109/JPROC.2021.3052449.

[3] L. Bergman and Yedid Hoshen, "Classification-Based Anomaly Detection for General Data," *Openreview.net*, 2020. https://openreview.net/forum?id=H1lK_lBtvS (accessed May 08, 2026).

[4] P. Malhotra, A. Ramakrishnan, G. Anand, L. Vig, P. Agarwal, and G. R. Shroff, "LSTM-based Encoder-Decoder for Multi-sensor Anomaly Detection," Jul. 2016, doi: https://doi.org/10.48550/arxiv.1607.00148.

[5] T. Zonta, C. A. da Costa, R. da Rosa Righi, M. J. de Lima, E. S. da Trindade, and G. P. Li, "Predictive maintenance in the Industry 4.0: A systematic literature review," Computers & Industrial Engineering, vol. 150, p. 106889, Dec. 2020, doi: https://doi.org/10.1016/j.cie.2020.106889

[6] A. Iqbal, R. Amin, F. S. Alsubaei, and A. Alzahrani, "Anomaly detection in multivariate time series data using deep ensemble models," PLOS ONE, vol. 19, no. 6, p. e0303890, Jun. 2024, doi: https://doi.org/10.1371/journal.pone.0303890.

[7] M. Ahmed, A. Naser Mahmood, and J. Hu, "A survey of network anomaly detection techniques," *Journal of Network and Computer Applications*, vol. 60, pp. 19–31, Jan. 2016, doi: https://doi.org/10.1016/j.jnca.2015.11.016.

[8] "Goodfellow, I., et al. (2016) Deep Learning. MIT Press, Cambridge, MA. - References - Scientific Research Publishing," *www.scirp.org*. https://www.scirp.org/reference/referencespapers?

[9] K. Zhang et al., "Self-Supervised Learning for Time Series Analysis: Taxonomy, Progress, and Prospects," IEEE transactions on pattern analysis and machine intelligence, pp. 1–20, Jan. 2024, doi: https://doi.org/10.1109/tpami.2024.3387317.

[10] T. P. Carvalho, F. A. A. M. N. Soares, R. Vita, R. da P. Francisco, J. P. Basto, and S. G. S. Alcalá, "A systematic literature review of machine learning methods applied to predictive maintenance," *Computers & Industrial Engineering*, vol. 137, no. 1, p. 106024, Nov. 2019, doi: https://doi.org/10.1016/j.cie.2019.106024.