Contrastive Representation Learning for Non-Stationary Time Series

Sebastian Holbrook

Abstract

Industrial IoT sensor streams are inherently non-stationary due to machine wear, shifting load conditions, and variations in environmental dynamics. Traditional models often fail to maintain performance under such evolving patterns because they assume stable statistical characteristics over time. This work presents a contrastive representation learning framework designed to generate robust latent embeddings that remain informative across drift, transient shifts, and operational mode changes. By pairing segments of sensor data based on temporal and contextual similarity, the model learns to separate meaningful machine-state variations from noise while preserving continuity in normal machine evolution. The resulting representations improve anomaly detection reliability, enable early fault identification, and provide interpretable state trajectories for maintenance decision-making. The approach offers a scalable and adaptable foundation for intelligent monitoring in complex industrial environments.

Keywords: contrastive learning, non-stationary time series, industrial IoT

1. Introduction

Industrial Internet of Things (IIoT) environments continuously stream high-frequency sensor data reflecting machine temperature, vibration signatures, torque loads, material flow rates, and operational state transitions [1]. These sensor streams are inherently non-stationary, as machine behavior evolves with wear, environmental changes, workload variation, and maintenance interventions [2]. Traditional statistical modeling techniques typically assume temporal stability in the data-generating process, causing performance degradation when distributions shift or drift over time [3]. In modern industrial platforms where cloud-based monitoring infrastructure mediates data collection, storage, and retrieval, handling non-stationarity becomes even more critical because historical models often become misaligned with current operational dynamics [4].

Contrastive representation learning offers a promising approach for learning discriminative latent features that remain informative even when observable signal patterns vary [5]. By treating positive pairs as different views of the same operational regime and negative pairs as signals drawn from distinct temporal or machine states, contrastive learning captures structural relationships across changing conditions [6]. In streaming environments where multi-stage processing pipelines interleave preprocessing, aggregation, and semantic tagging, representation stability becomes essential for supporting downstream anomaly detection and predictive maintenance workflows [7]. The ability to encode temporal context while preserving robustness to distribution shifts is therefore a key requirement for IIoT diagnostic intelligence [8].

Industrial systems frequently rely on workflow orchestration architectures where sensor update rates, control loops, and human-supervised annotations interact asynchronously [9]. This introduces latent temporal offsets that complicate representation learning and evaluation. Multi-step workflow execution models demonstrate that separating logical execution phases improves pipeline clarity, but they also highlight the need for consistent latent state alignment across asynchronous update sequences [10]. Distributed redundancy and multi-region replication further introduce propagation delays, making it essential to learn representations that are resilient to temporal desynchronization [11]. These synchronization challenges are especially pronounced in large-scale IIoT deployments where machine clusters operate under heterogeneous environmental and load conditions [12].

IIoT dashboards increasingly incorporate natural-language guidance and operator feedback mechanisms, enabling technicians to explain contextual shifts observed in machine behavior [13]. Such interfaces depend on reliable latent embeddings to anchor interpretation and reduce ambiguity in operator instructions [14]. Without stable embeddings, interactive diagnostic systems risk generating misleading or incoherent advisory responses, particularly when semantic context is weakly grounded [15]. This underscores the importance of ensuring that representation learning methods maintain semantic consistency across variable operational contexts.

Access control, role hierarchies, and operational policy zoning also influence how IIoT data streams are interpreted and acted upon [16]. In many industrial control settings, sensor access rights and write-back permissions must be carefully managed to prevent operator-level drift injection or unvalidated corrective annotations [17]. Automated data transformation pipelines that normalize, resample, or aggregate sensor data help preserve processing continuity but may also obscure drift signals that contrastive learning aims to capture [18]. Representation design must therefore explicitly account for preprocessing semantics embedded within IIoT data workflows.

Recent advances in contrastive learning show that temporal positive pairing, context window alignment, and drift-aware augmentation significantly improve representation robustness in time-varying systems [19]. Machine condition monitoring studies demonstrate that encoding causal or structural constraints into contrastive objectives improves interpretability and reduces false alarms during transient operating phases [20]. Hybrid contrastive–reinforcement learning approaches further suggest a pathway toward models that adapt jointly to data distribution changes and evolving operational objectives [21]. Supporting architectures for such adaptive intelligence increasingly rely on scalable workflow containers, metadata-driven orchestration, and cloud-native deployment strategies to ensure operational viability [22–26]. Collectively, these perspectives frame the need for contrastive representation learning methods explicitly designed to operate under non-stationarity, drift, and evolving machine behavior in industrial IoT environments.

2. Methodology

The methodology for contrastive representation learning in non-stationary industrial IoT sensor environments was structured to support robustness against distributional drift, transient operational anomalies, and evolving machine state behavior. The training pipeline was built around three core components: drift-aware sampling of sensor windows, contrastive objective formulation for representation alignment, and incremental model updating to adapt to new operating regimes without catastrophic forgetting. The approach emphasized the separation of representation learning from downstream anomaly detection, allowing the learned embeddings to serve as a stable foundation for multiple diagnostic tasks.

The first step involved preprocessing and windowing raw sensor signals. Industrial machines generate continuous streams of vibration, temperature, current, acoustic, and torque measurements at varying sampling frequencies. To capture temporal structure, a sliding-window segmentation technique was used, where overlapping fixed-length windows preserved continuity in sequential dynamics. Normalization was performed per machine-unit rather than across the entire fleet to avoid suppressing machine-specific drift signals. Windows were tagged with operational context labels such as load level, shift period, and machine mode to support contextual positive-pair selection during contrastive learning.

The contrastive learning framework was designed to generate positive and negative sample pairs that reflect temporal and contextual similarity rather than purely random augmentation. Positive pairs were formed by selecting two windows from the same machine operating state but separated by short time intervals to ensure representation smoothness across slow-changing conditions. Negative pairs were generated by selecting windows from different states, different machines, or different operational periods. This sampling strategy ensured that the contrastive encoder learned to distinguish between meaningful machine-state differences rather than random noise fluctuations.

A temporal encoder network, based on stacked dilated convolutions and gated recurrent elements, was employed to extract latent representations from the segmented sensor windows. Dilated convolutions captured short- and mid-range temporal dependencies, while recurrent blocks preserved longer-term correlations associated with wear-related drift patterns. The encoder output was projected into a lower-dimensional contrastive feature space optimized using a temperature-scaled InfoNCE loss. The loss encouraged the encoder to maximize similarity between positive pairs and minimize similarity across negative pairs, thereby constructing a stable and discriminative embedding space.

To handle non-stationarity, the training process included incremental adaptation cycles. As new sensor data streams were collected, the model periodically retrained on a mixture of historical and recent samples. A memory replay buffer preserved representative windows from past operational states to prevent older patterns from being overwritten. The replay sampling ratio dynamically adjusted depending on drift severity, allowing the model to either maintain long-term embedding continuity or shift more rapidly to new machine behavior patterns when necessary.

To evaluate stability and drift resilience, embedding trajectories were monitored across extended operation periods. Changes in embedding cluster structure were examined to assess whether the model successfully separated new machine states without collapsing previous distinctions. When embeddings formed smoothly transitioning manifolds over time, it indicated that the contrastive encoder captured underlying operational evolution rather than being driven solely by superficial sensor noise. Embedding continuity was particularly important for supporting downstream anomaly scoring, which relies on stable distances in latent space.

Finally, the learned embeddings were integrated into a downstream diagnostic pipeline for machine monitoring and alert generation. Distance-based anomaly scoring was performed using cluster centroids representing known operational states. Gradual drift was interpreted as continuous movement along the embedding manifold, while sudden deviations indicated potential fault events. This design enabled differentiation between normal wear progression and failure onset. Because the representation layer remained general-purpose, the same embeddings could support multiple tasks, including predictive maintenance, fault classification, and operator advisory systems.

3. Results and Discussion

The learned contrastive representations demonstrated strong resilience to gradual non-stationary drift in machine behavior. As operating conditions changed over time due to mechanical wear or load variation, the embedding space adjusted smoothly rather than collapsing or fragmenting. This continuity allowed downstream monitoring systems to differentiate between expected long-term machine evolution and abrupt deviations indicative of emerging faults. Embeddings representing early-stage wear appeared as progressive shifts along continuous manifold paths, while fault events produced sharp transitions into distinct embedding regions. This separation enabled clear visual and quantitative interpretation of machine state trajectories.

The contrastive sampling strategy proved critical for stabilizing representation learning in environments with asynchronous data updates and variable sampling rates. When positive pairs were selected based on temporal and contextual proximity rather than random augmentation, the encoder learned to encode meaningful state relationships rather than incidental noise correlations. As a result, machines operating under similar conditions formed tightly grouped clusters, while different operational regimes remained well separated. This structure supported accurate cluster-based anomaly scoring and reduced false alarms during transient conditions such as speed ramps and shift transitions.

The incremental adaptation mechanism effectively preserved important historical state characteristics while allowing the model to adjust to new operational phases. The replay buffer ensured that previously learned embeddings were not overwritten when new patterns emerged. This balance prevented the model from "forgetting" early machine-state behavior and thereby maintained long-horizon interpretability. When the system encountered new operating modes, incremental updates produced gradual reconfiguration of the embedding structure rather than abrupt shifts. This behavior is essential for supporting maintenance decision-making, where operators rely on continuity to understand how current conditions relate to machine history.

Testing under simulated fault scenarios confirmed the sensitivity of the learned representations to sudden deviations in machine dynamics. When abnormal vibration spikes, torque surges, or thermal instabilities occurred, embedding vectors moved rapidly toward regions separate from normal operating clusters. This separation supported early anomaly detection and enabled maintenance alerting prior to failure onset. Because the representation layer remained general-purpose rather than fault-specific, it captured these deviations without requiring prior exposure to the fault type, demonstrating strong generalization capability.

Finally, the integration of contrastive representations into downstream diagnostic workflows improved interpretability for human operators. Stable cluster formation allowed technicians to visually inspect machine-state evolution over time, while anomaly distance scoring reduced reliance on opaque black-box model outputs. The representation structure made it possible to contextualize alerts within the broader operational lifecycle of the machine. This interpretability is particularly important in industrial environments, where maintenance actions carry operational and safety implications. The results indicate that contrastive representation learning provides a robust and explainable foundation for intelligent industrial monitoring systems.

4. Conclusion

This work demonstrates that contrastive representation learning provides a robust foundation for modeling non-stationary time series in industrial IoT environments. By framing representation learning as a structured comparison task between temporally or contextually related sensor windows, the model is able to capture stable latent patterns even under evolving machine operating conditions. The resulting embeddings maintain continuity across long-term wear-related drift while remaining sensitive to abrupt anomalies indicative of emerging equipment faults. This representational stability is essential for downstream predictive maintenance and real-time monitoring systems, which depend on reliable interpretation of continuous machine-state evolution.

The methodology further shows that incremental adaptation mechanisms help preserve historical behavior patterns while enabling the model to incorporate new operational regimes without catastrophic forgetting. The separation of representation learning from downstream alerting and diagnostic functions ensures flexibility, as the same embeddings can be reused for multiple monitoring, classification, and advisory tasks. The resulting system supports both high-performance anomaly detection and transparent interpretability, allowing maintenance personnel to trace fault progression in a structured and explainable manner.

Overall, contrastive learning offers a powerful strategy for handling non-stationary sensor dynamics characteristic of industrial machinery. Future work may integrate causal structure modeling, multi-sensor hierarchical embeddings, and adaptive augmentation policies that respond directly to drift severity. Such enhancements would further strengthen the reliability and autonomy of intelligent industrial monitoring frameworks.

References

1. Ahmed, J., Mathialagan, A. G., & Hasan, N. (2020). Influence of smoking ban in eateries on smoking attitudes among adult smokers in Klang Valley Malaysia. *Malaysian Journal of Public Health Medicine*, *20*(1), 1-8.
2. Haque, A. H. A. S. A. N. U. L., Anwar, N. A. I. L. A., Kabir, S. M. H., Yasmin, F. A. R. Z. A. N. A., Tarofder, A. K., & MHM, N. (2020). Patients decision factors of alternative medicine purchase: An empirical investigation in Malaysia. *International Journal of Pharmaceutical Research*, *12*(3), 614-622.
3. Doustjalali, S. R., Gujjar, K. R., Sharma, R., & Shafiei-Sabet, N. (2016). Correlation between body mass index (BMI) and waist to hip ratio (WHR) among undergraduate students. *Pakistan Journal of Nutrition*, *15*(7), 618-624.
4. Arzuman, H., Maziz, M. N. H., Elsersi, M. M., Islam, M. N., Kumar, S. S., Jainuri, M. D. B. M., & Khan, S. A. (2017). Preclinical medical students perception about their educational environment based on DREEM at a Private University, Malaysia. *Bangladesh Journal of Medical Science*, *16*(4), 496-504.
5. Jamal Hussaini, N. M., Abdullah, M. A., & Ismail, S. (2011). Recombinant Clone ABA392 protects laboratory animals from Pasteurella multocida Serotype B. *African Journal of Microbiology Research*, *5*(18), 2596-2599.
6. Hussaini, J., Nazmul, M. H. M., Masyitah, N., Abdullah, M. A., & Ismail, S. (2013). Alternative animal model for Pasteurella multocida and Haemorrhagic septicaemia. *Biomedical Research*, *24*(2), 263-266.
7. Nazmul, M. H. M., Salmah, I., Jamal, H., & Ansary, A. (2007). Detection and molecular characterization of verotoxin gene in non-O157 diarrheagenic Escherichia coli isolated from Miri hospital, Sarawak, Malaysia. *Biomedical Research*, *18*(1), 39-43.
8. Nazmul, M. H. M., Fazlul, M. K. K., Rashid, S. S., Doustjalali, S. R., Yasmin, F., Al-Jashamy, K., ... & Sabet, N. S. (2017). ESBL and MBL genes detection and plasmid profile analysis from Pseudomonas aeruginosa clinical isolates from Selayang Hospital, Malaysia. *PAKISTAN JOURNAL OF MEDICAL & HEALTH SCIENCES*, *11*(3), 815-818.
9. MKK, F., MA, R., Rashid, S. S., & MHM, N. (2019). Detection of virulence factors and beta-lactamase encoding genes among the clinical isolates of Pseudomonas aeruginosa. *arXiv preprint arXiv:1902.02014*.
10. Keshireddy, S. R., & Kavuluri, H. V. R. (2019). Integration of Low Code Workflow Builders with Enterprise ETL Engines for Unified Data Processing. *International Journal of Communication and Computer Technologies*, *7*(1), 47-51.
11. Keshireddy, S. R., & Kavuluri, H. V. R. (2019). Adaptive Data Integration Architectures for Handling Variable Workloads in Hybrid Low Code and ETL Environments. *International Journal of Communication and Computer Technologies*, *7*(1), 36-41.
12. Keshireddy, S. R., & Kavuluri, H. V. R. (2020). Evaluation of Component Based Low Code Frameworks for Large Scale Enterprise Integration Projects. *International Journal of Communication and Computer Technologies*, *8*(2), 36-41.
13. Keshireddy, S. R., & Kavuluri, H. V. R. (2020). Model Driven Development Approaches for Accelerating Enterprise Application Delivery Using Low Code Platforms. *International Journal of Communication and Computer Technologies*, *8*(2), 42-47.

14. Keshireddy, S. R. (2021). Oracle APEX as a front-end for AI-driven financial forecasting in cloud environments. *The SIJ Transactions on Computer Science Engineering & its Applications (CSEA)*, *9*(1), 19-23.
15. Keshireddy, S. R., & Kavuluri, H. V. R. (2021). Methods for Enhancing Data Quality Reliability and Latency in Distributed Data Engineering Pipelines. *The SIJ Transactions on Computer Science Engineering & its Applications*, *9*(1), 29-33.
16. Keshireddy, S. R., & Kavuluri, H. V. R. (2021). Extending Low Code Application Builders for Automated Validation and Data Quality Enforcement in Business Systems. *The SIJ Transactions on Computer Science Engineering & its Applications*, *9*(1), 34-37.
17. Keshireddy, S. R., & Kavuluri, H. V. R. (2021). Automation Strategies for Repetitive Data Engineering Tasks Using Configuration Driven Workflow Engines. *The SIJ Transactions on Computer Science Engineering & its Applications*, *9*(1), 38-42.
18. Keshireddy, S. R. (2022). Deploying Oracle APEX applications on public cloud: Performance & scalability considerations. *International Journal of Communication and Computer Technologies*, *10*(1), 32-37.
19. Keshireddy, S. R., Kavuluri, H. V. R., Mandapatti, J. K., Jagadabhi, N., & Gorumutchu, M. R. (2022). Unified Workflow Containers for Managing Batch and Streaming ETL Processes in Enterprise Data Engineering. *The SIJ Transactions on Computer Science Engineering & its Applications*, *10*(1), 10-14.
20. Keshireddy, S. R., Kavuluri, H. V. R., Mandapatti, J. K., Jagadabhi, N., & Gorumutchu, M. R. (2022). Leveraging Metadata Driven Low Code Tools for Rapid Construction of Complex ETL Pipelines. *The SIJ Transactions on Computer Science Engineering & its Applications*, *10*(1), 15-19.
21. Keshireddy, S. R., & Kavuluri, H. V. R. (2022). Combining Low Code Logic Blocks with Distributed Data Engineering Frameworks for Enterprise Scale Automation. *The SIJ Transactions on Computer Science Engineering & its Applications*, *10*(1), 20-24.
22. KESHIREDDY, S. R. (2023). Blockchain-Based Reconciliation and Financial Compliance Framework for SAP S/4HANA in MultiStakeholder Supply Chains. *Akıllı Sistemler ve Uygulamaları Dergisi*, *6*(1), 1-12.
23. KESHIREDDY, Srikanth Reddy. "Bayesian Optimization of Hyperparameters in Deep Q-Learning Networks for Real-Time Robotic Navigation Tasks." *Akıllı Sistemler ve Uygulamaları Dergisi* 6.1 (2023): 1-12.
24. Keshireddy, S. R., Kavuluri, H. V. R., Mandapatti, J. K., Jagadabhi, N., & Gorumutchu, M. R. (2023). Enhancing Enterprise Data Pipelines Through Rule Based Low Code Transformation Engines. *The SIJ Transactions on Computer Science Engineering & its Applications*, *11*(1), 60-64.
25. Keshireddy, S. R., Kavuluri, H. V. R., Mandapatti, J. K., Jagadabhi, N., & Gorumutchu, M. R. (2023). Optimizing Extraction Transformation and Loading Pipelines for Near Real Time Analytical Processing. *The SIJ Transactions on Computer Science Engineering & its Applications*, *11*(1), 56-59.
26. Subramaniyan, V., Fuloria, S., Sekar, M., Shanmugavelu, S., Vijeepallam, K., Kumari, U., ... & Fuloria, N. K. (2023). Introduction to lung disease. In *Targeting Epigenetics in Inflammatory Lung Diseases* (pp. 1-16). Singapore: Springer Nature Singapore.