Volume 2, 2026, Pages 252-264 *Elaris Computing Nexus*
| Regular Article | Open Access

Topology Aware Temporal Contrastive Learning Framework for Robust Multivariate IoT Cyber Physical System Monitoring

Arulkumar Natarajan
Department of Computer Science, Samarkand International University of Technology, Samarkand, Uzbekistan.
arul.natarajan@siut.uz

Article Info
Elaris Computing Nexus
https://elarispublications.com/journals/ecn/ecn_home.html

Received 18 June 2026
Revised from 26 August 2026
Accepted 24 September 2026
Available online 06 November 2026
Published by Elaris Publications

©The Author(s), 2026.
https://doi.org/10.65148/ECN/2026021

Corresponding author(s):
Arulkumar Natarajan, Samarkand International University of Technology, Samarkand, Uzbekistan.
Email: arul.natarajan@siut.uz

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Abstract - Cyber-Physical Systems (CPS) have been increasingly deployed in the world of Internet of Things (IoT) which results in a continuous multivariate time-series data with fluctuating operational patterns, nonlinear dynamics and temporal dependencies. To be useful for monitoring, representations of such systems must accurately model the temporal nature of sensor signals and the relationships between them. The problem of robust monitoring in IoT-CPS is tackled in this study by using a topology-aware temporal contrastive learning framework, namely TopoTime-CPS. The proposed framework is based on a temporal feature extraction and a topological feature construction which is used to represent the dynamic pattern of the signal and the geometric structure of the multivariate sensor observations. A delay-embedding method is employed to transform the time-series measurements into state space and a topology-preserving method was used to extract features using variation-regulated filtration and homology analysis. Meanwhile, for temporal representation, the temporal discrimination of normal and abnormal operating patterns is enhanced by employing hierarchical contrastive learning. This results in a temporal and topological representation, then mapped to a common latent representation and aligned by a time-topology correspondence objective. The common representation helps to preserve complementary temporal and structural information, and a decreased sensitivity to transient variations of the sensors. The framework can be adapted to monitoring of IoT-CPS and used as a downstream anomaly detection and operational-state recognition. The proposed approach will be evaluated in an experimental setting using representative Machine Learning (ML) and Deep Learning (DL) approaches using quantitative performance metrics and ablation studies.

Keywords - Internet of Things, Cyber-Physical Systems, Temporal Contrastive Learning, Topological Data Analysis, Multivariate Time Series, Anomaly Detection.

I. INTRODUCTION AND RESEARCH CONTEXT

With the evolvement of the IoT, CPS can now continuously gather information from the interconnected sensors, actuators, machines and communication devices. They create multivariate time-series data with different sampling rates and operating conditions for various applications such as industrial automation, smart buildings, energy systems, transportation, environmental monitoring, and healthcare. Each of the collected measurements have useful information about the state of the system and can be utilized during anomaly detection, fault diagnosis, predictive maintenance and operation. The ability to reliably and discriminatively extract meaningful and useful representations from these types of data is challenging due to the dynamism, correlation, nonlinearity and perhaps noisiness and transience of CPS signals [1]. Most machine-learning techniques are designed in a statistical and/or frequency domain and effort-intensive. These methods may function well in a controlled environment, but may not be as effective in changing operating conditions or if the abnormal patterns are not pronounced. Hence, deep-learning techniques are promising as they can learn to represent the data hierarchically without manual coding, directly from the measurements taken by the sensors. Local and long-range dependency in the sequences of sensors can be captured by temporal models like convolutional networks, recurrent networks, and attention-based architectures. But just temporally representing relationships between observations is not necessarily structuring their

relationships. Two segments of a time-series can have differing short-term patterns, which are in significantly different system states [2].

Topological data analysis (TDA) offers an alternative view, which is based on the structure of the data instead of just the values of the signals. A one-dimensional sequence of sensors may be embedded to get a higher dimensional state space representation that is delay-embedded. This representation allows to describe topological structures by filtering and homology. These descriptors can retain connectivities, loops and other geometric properties that can vary with an alteration in the physical process. This means that the topological information could be utilized in abnormal normal recognition for CPS [3]. Despite all of these advantages, independent representation learning in time and space can lead to the emergence of non-uniform feature spaces. These representations will not necessarily result in a semantically matched temporal and structural pattern if a direct match occurs in such a relationship. Additionally, CPS monitoring needs to involve representations which are informative under the presence of sensor noise and/or changing operating conditions. With the close relation between semantically similar representations and the repulsion between semantically dissimilar representations in contrastive learning, this problem can be addressed properly. While conventional TCL considers the external dynamics of the sensors, it does not explicitly take into account the topological features of the sensors' dynamics [4].

This inspires this work, which proposes a topology-aware temporal contrastive learning framework for monitoring IoT-CPS (TopoTime-CPS). The structure learns a topological representation and temporal representation independently, and learns the conversion from the temporal representation to the topological representation in a shared latent representation space. To encode complementary properties at various temporal levels [5], the branch is designed as a hierarchical contrastive learning branch called temporal branch. Topological branch: Modeling and construction of features that capture the structural nature of the sensor dynamics from the delay-embedded time series and variation-regulated filter / homology. These learned representations are then projected onto a similar feature space for a time-topology correspondence objective, which enforces the same system state to be represented in the same way. Finally, the resulting representation will be used to fuse both time dynamics and structure data aiming at avoiding the construction of an excessively complicated end-to-end architecture [6].

Motivation of the Research

The primary motivation is to achieve a powerful representation that can communicate on how things change with time and the organisation of the underlying sensor dynamics. Temporal-only representations might miss geometric changes while topology-only representations might miss important sequential information. Thus, it is important to consider the complementary aspects and adopt contrastive representation learning to integrate the two views for IoT-CPS monitoring. The proposed framework specifically focuses on providing a better representation, robustness against signal variations and discrimination between various operating conditions while keeping the architecture small enough.

Research Contributions and Novelty

The novelty of this research is the use of a time-topology aligned contrastive representation framework, instead of concatenating temporal and topological features. Three novel ideas are introduced: (i) multi-level temporal contrastive representation learning for temporal information at multiple levels, (ii) constructing features designed by filter and homology, considering the delay of sensor dynamics, and (iii) a time-topology correspondence objective that aligns the two representations in a common latent space. The purpose of this design is to capture a unified view of monitoring for IoT-CPS which could enable systematic evaluation via comparative experiments, ablation analysis, robustness evaluation and latent space visualization.

Organization of the Article

This article is divided into several parts as follows: In Section 2, the related works in IoT-CPS monitoring, temporal representation learning, contrastive learning and topological data analysis are introduced. The proposed framework for TopoTime-CPS and its key components is described in Section 3. Section 4 is about experimental setup, experimental datasets, evaluation metrics, and implementation configuration. Section 5 presents the quantitative and qualitative experimental results, including comparative, ablation and robustness analysis. Finally, the study is concluded and suggestions are made for future research in Section 6.

II. RELATED STUDIES AND RESEARCH GAP

Multivariate sensor streams are constantly created in IoT and cyber-physical systems where abnormal behavior can be caused by equipment faults, environmental disturbances, communication issues, or cyber-physical attacks. Therefore, multivariate time-series anomaly detection is ongoing research for the reliable monitoring of the CPS is active. The current methods can be classified into three types, e.g., forecasting, reconstruction and contrastive-learning, and modelling temporal dependencies and relationship between multiple variables is difficult [7]. Most initial methods were based on statistical models, distance-based methods, and classical machine-learning methods. They are generally effective with hand-picked features and assumptions regarding the distribution of signals, but have been found to be computationally attractive. New nonlinear temporal aspects were automatically learned by deep autoencoder and recurrent architectures

which improved representation learning further. However, reconstruction algorithms can have difficulties in cases where the distance between the normal and abnormal patterns in the data space is very small or when anomalies occur in multiple dimensions of the data space [8].

A way to resolve this, in part, was to explicitly model inter-sensor relationships in the form of graphs. For example, Graph Deviation Network (GDN) can learn the relationship between various sensors and find the deviations in the relationship between sensors via graph attention forecasting. Likewise, GTAD incorporates temporal convolution and graph attention into a unified framework, which is able to capture temporal dependency and sensor dependency. These methods show that taking into account the relationships between the sensors is important, and not just focusing on each sensor's measurement individually. For most of the time however, the pairwise relationships are represented in a graph rather than a full description of the geometry in the dynamics of a CPS [9].

More improvements have been made by transformer models for the modelling of long-range temporal dependency. Transformer-based sequence encoding, self-conditioning and adversarial training are used to perform multivariate anomaly detection and diagnosis in TranAD. There are other approaches, based on dynamic graph-transformers, which have tried to model temporal and inter-sensor dependence jointly, showing the need for dynamic, rather than static, dependence. These methods are more strongly based on taught temporal or graph structures, and fail to explicitly retain the topological structure of the underlying signal trajectories [10]. Another direction that is promising is self-supervised and contrastive representation learning, as the availability of labeled CPS anomaly data is often limited. Given the case of few labels in time-series, Temporal Neighborhood Coding (TNC) seeks to learn representation using temporal neighbourhoods and contrastive objectives. TS2Vec extends this approach with hierarchical contrastive learning in a long-term temporal context, and provides multi-level representation. These methods are able to learn very informative temporal representations without vast amounts of anomaly labels. However, most of their contrastive goals are in time representation space, and they do not explicitly feature correspondence to topological properties [11].

The TDA describes the geometric structuring of time-series trajectories in a complementary way, by offering another way of describing it. Again, embedding can be done to represent a scalar temporal sequence as a state space representation and then filtering and homology may be adopted to describe its structural properties, which may not be well captured by the classical temporal encoders. Recently, researchers have begun looking at the problem of multivariate anomaly detection from a perspective of persistent homology with potential for geometric information in Iot which is noise-invariant. Persistent homology inspired graph-based models have also have been discussed recently for tackling the problem of anomaly detection in Industrial IoT (IIoT) systems, indicating a growing interest in learning topology representations. In recent years, persistent homology inspired graph-based models have also been discussed for tackling the problem of anomaly detection in Industrial IoT (IIoT) systems, which has marked an increasing interest in learning topology representations. Even with the advancement accomplished there is a great void that remains to be bridged. Most of the previous temporal contrastive methods are based on temporal invariance and topology-based methods are mostly based on independent structural descriptors based on topology. Graph-based methods explicitly describe the dependence of any two sensors and learned methods but they ignore the explicitly representation of higher order topological features. Therefore, the common representation-learning mechanism, which is able to learn two hierarchical temporal properties and explicitly map the topological structure, has not yet been thoroughly explored to be applied to monitoring of IoT-CPS. This strongly supports the proposed framework, instead of merely merging heterogeneous features together, we propose to include a time-topology correspondence objective.

Table 1 summarizes properties of the representative approaches mentioned in multivariate IoT-CPS monitoring: conventional machine learning, graph based, Transformer based, temporal contrastive, and topology-oriented. Through the comparison, it is observed that, at a limited extent, hierarchical temporal representation, topological structure and explicit cross-representation alignment were integrated. In this comparison, it is noted that integration of hierarchical temporal representation, topological structure and explicit cross-representation alignment has been limited, which motivates the proposed TopoTime-CPS framework. This is the goal of the proposed TopoTime-CPS framework for learning a common representation for multivariate sensor data in IoT-CPS with the exploitation of temporal dynamics and topological structure of the sensor trajectories. It involves three steps: (i) Temporal Representation Learning: Representing local and global temporal patterns in normalized temporal sequences of sensors, (ii) Topological Structure Encoding: Embedding each sensor observation in a state space that is fixed and has a topology and can compute (structural) descriptors that are persistent over time, and (iii) Time-Topology Contrastive Alignment: The two representations are brought into a common latent space, and the maximality of the alignment is achieved using a contrastive objective function. The aligned representation is then fed to the CPS state characterization and decision making based on anomaly. The proposed framework makes it explicit that the same operating condition can be represented both topologically and temporally in different ways, which may become semantically consistent, and separates representations of the different system states.

Temporal Representation Learning

The first stage of TopoTime-CPS converts the normalized multivariate sensor sequence into a compact temporal representation. Let the input sequence be represented as ($\boldsymbol{X} = \{\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, \boldsymbol{x_T}\}$), from ($d$) IoT sensors at time ($t$), and ($T$) is the observation length. The values obtained from the sensors are first normalized so that they can be used on sensors that

have different scales. Then, a temporal encoder is tasked with extracting short-term variations, temporal dependencies, and evolving operating patterns from overlapping windows of the normalized sequence.

Table 1. Comparative Analysis of Representative Approaches for IoT-CPS Multivariate Time-Series Monitoring

Method / Category	Temporal Features	Inter-Sensor Relations	Topological Information	Contrastive Learning	Multi-Level Representation	Main Limitation
Statistical / Machine Learning Methods [9]	Statistical and handcrafted temporal features	Explicit feature correlations	Not considered	Not used	Single-level features	Dependence on handcrafted features and distribution assumptions
Recurrent / Autoencoder Methods [13]	Sequential and nonlinear temporal features	Implicitly learned	Not considered	Not used	Limited hierarchical representation	Reconstruction may fail to distinguish subtle anomalies
GDN [14]	Temporal forecasting features	Graph-based sensor dependencies	Not considered	Not used	Limited multi-scale representation	Mainly captures pairwise sensor relationships
GTAD [14]	Temporal convolutional features	Graph-attention-based relationships	Not considered	Not used	Multi-level temporal representation	Higher-order geometric structure is not explicitly modeled
TranAD [15]	Long-range Transformer-based temporal features	Implicit sensor dependencies	Not considered	Not used	Hierarchical sequence representation	Focuses primarily on temporal dependencies
TNC [16]	Local temporal neighborhood features	Limited relationship modeling	Not considered	Temporal neighborhood contrastive learning	Multi-scale temporal representation	Does not explicitly model topological structure
TS2Vec [16]	Local and global temporal features	Limited inter-variable modeling	Not considered	Hierarchical temporal contrastive learning	Hierarchical representation	Temporal representation is not explicitly aligned with topology
TDA-Based Methods	Conventional or limited temporal features	Indirect relationships	Persistent homology and topological descriptors	Generally, not used	Limited structural representation	Topological and temporal representations are commonly learned separately
Topology-Enhanced Graph Methods	Temporal and graph-based features	Learned graph relationships	Persistent topological descriptors	Limited contrastive objectives	Joint graph-topological representation	Explicit temporal-topological correspondence remains limited
TopoTime-CPS (Proposed)	Hierarchical temporal representation	Learned sensor relationships	Delay embedding, filtration and homology	Hierarchical temporal and cross-representation contrastive learning	Joint temporal-topological representation	Requires construction of both temporal and topological feature branches

To capture the information in multiple temporal scales, the encoder learns representations at various temporal resolutions so that the rapid temporal changes in the signal are retained in the learned feature space, while slower system-level dynamics remain. The temporal representation generated is the first complementary view of the CPS state, and is then mapped to the topological representation in the proposed contrastive learning stage.

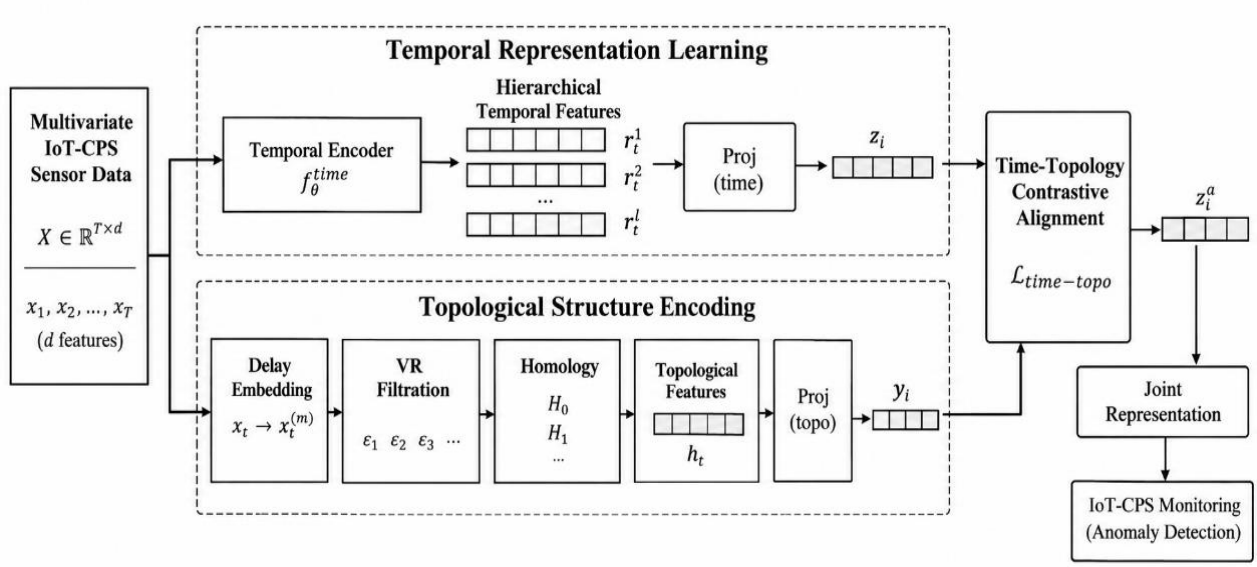

Fig 1. Proposed TopoTime-CPS Framework for Time-Topology Contrastive Learning in IoT-CPS Monitoring.

The overall architecture of the proposed TopoTime-CPS architecture is shown in **Fig. 1**. The framework uses two complementary processing branches to learn both temporal and topological representations given a multivariate IoT-CPS sensor sequence (X_i). The temporal branch uses a temporal encoder to learn hierarchical temporal dependencies, while the topological branch builds delay-embedded representations and uses variation-regulated filtering and homology analysis to learn structure. Using different projection functions, the resulting features are remapped into a common latent space. Finally, the proposed time-topology contrastive learning mechanism is proposed to align the projected temporal representation (z_i) and topological representation (y_i). The aligned representation gives a joint characterization of the underlying CPS operating condition in a compact format, which is then applied for monitoring based on anomalies.
The temporal and topological representations are formulated as

$$z_i = P_{\text{time}}\left(f_\theta^{\text{time}}(X_i)\right), \qquad y_i = P_{\text{topo}}\left(f_\theta^{\text{topo}}(X_i)\right) \tag{1}$$

where (f_θ^{time}) and (f_θ^{topo}) denote the temporal and topological representation functions, respectively, and (P_{time}) and (P_{topo}) denote the corresponding projection functions. The resulting representations (z_i) and (y_i) are subsequently optimized through the proposed time-topology correspondence objective.

Topological Structure Encoding
In the second stage, the evolution of the sensor measurements in the IoT-CPS are interpreted in terms of their structure. One-dimensional sensor sequence can include geometric patterns, which is hard to describe only with the traditional temporal features. As a result, each input segment is represented as a state-space delay-embedded representation with the temporally related observations mapped to a higher dimension. A variation regulated filtration is subsequently applied to progressively build the simplicial complexes from the embedded points and homology analysis is used to characterize the respective characteristics of the constructed structures. Topological descriptors are then extracted, and these are then converted into a fixed dimensional feature vector (h_i) to represent the complementary operating state for the CPS. This branch is intended to maintain structural variations of the dynamics of the sensors and to minimize the reliance on specific sensor signal amplitudes.

For a sensor sequence $(x_i(t))$, the delay-embedded state vector is defined as

$$\boldsymbol{x}_{i,t}^{(m)} = [x_i(t),\, x_i(t-\tau),\, \ldots,\, x_i(t-(m-1)\tau)] \tag{2}$$

where (m) denotes the embedding dimension and (τ) represents the time delay. The collection of embedded vectors forms the state-space trajectory used for subsequent topological analysis.

Time-Topology Contrastive Alignment

TopoTime-CPS is further developed to find a common representation between the temporal and topological feature spaces in the final stage. The two branches of the observation of the CPS may be equally strong, but the features learned from them may be different in terms of distribution and semantic structure. To solve this, both temporal representation (z_i) and topological representation (y_i) are jointly optimized and the joint optimization objective is a time-topology contrastive method. If a representation comes from the same sensor segment, this is considered a positive pair; if it comes from a different sensor segment, this is considered a negative pair. This forces the model to hold the common information that the two views share in the common operating-state, while hiding some variations in representations. The learned embedding thus describes the behavior of CPS in a more consistent manner than either the temporal or topological features taken individually. The similarity between a temporal representation and a topological representation is measured using cosine similarity:

$$s(z_i, y_j) = \frac{z_i^{\top} y_j}{|z_i|_2 \, |y_j|_2} \tag{3}$$

For a mini-batch containing (N) samples, the time-topology correspondence loss is formulated using a temperature-scaled contrastive objective:

$$\mathcal{L}_{time-topo} = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{\exp(s(z_i, y_i)/\tau_c)}{\sum_{j=1}^{N} \exp(s(z_i, y_j)/\tau_c)} \tag{4}$$

where (τ_c) denotes the temperature parameter controlling the concentration of the similarity distribution. Minimizing (4) increases the similarity between corresponding temporal-topological representations while reducing the similarity to non-corresponding samples. The final aligned representation is obtained by combining the two projected views:

$$z_i^a = \phi(z_i, y_i) \tag{5}$$

where $(\phi(\cdot))$ denotes the representation integration operation. The resulting (z_i^a) is subsequently used as the unified feature representation for IoT-CPS monitoring and anomaly detection.

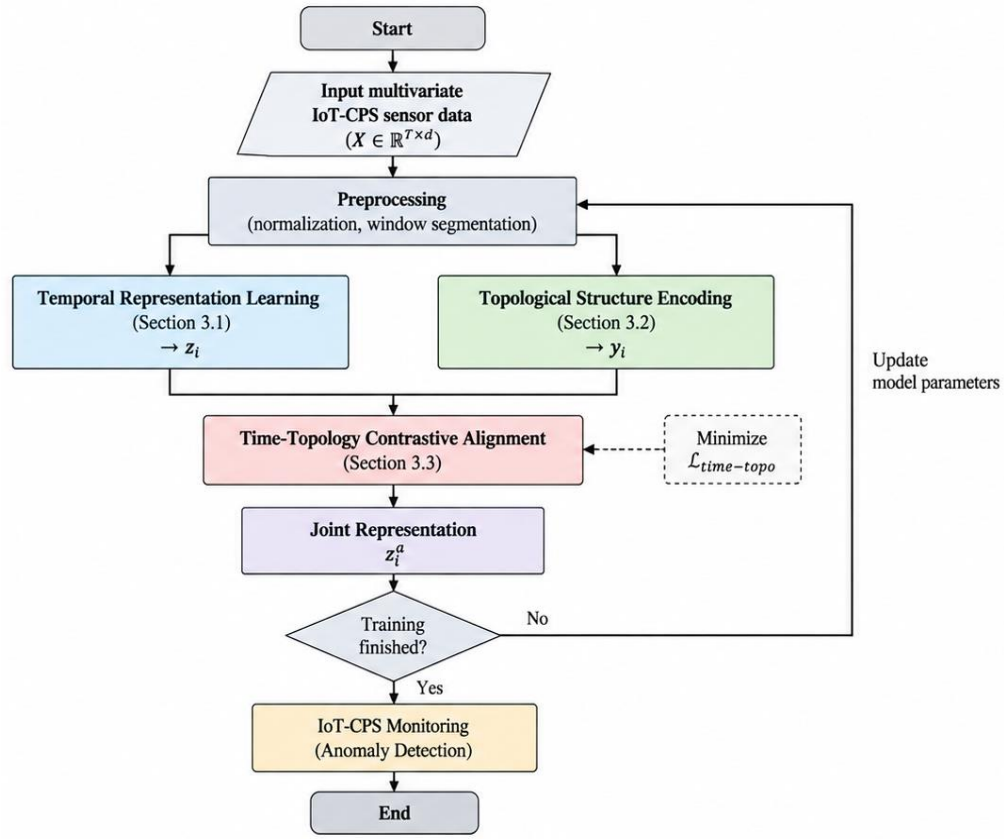

Fig 2. Flowchart of the Proposed TopoTime-CPS Framework.

Fig. 2 illustrates the computational workflow of the proposed TopoTime-CPS framework, beginning with multivariate IoT-CPS sensor data and preprocessing through normalization and window segmentation. The processed data are

simultaneously passed to the temporal representation learning and topological structure encoding stages. Their outputs are subsequently optimized through time-topology contrastive alignment to obtain the joint representation (z_i^a). During training, the proposed objective ($\mathcal{L}_{time-topo}$) is minimized and the model parameters are iteratively updated until the training criterion is satisfied. The learned joint representation is finally used for IoT-CPS anomaly detection.

III. EXPERIMENTAL SETUP AND EVALUATION PROTOCOL

The experimental setup to confirm the proposed TopoTime-CPS framework for monitoring an IoT-CPS is described in this section. The evaluation aims to test the effectiveness of the joint temporal-topological representation for improved discrimination of normal and anomalous operating conditions by compared with existing temporal, graph-based, and representation-learning methods. The protocol for experimental procedure includes the data set preparation, implementation configurations, evaluation measures, and baseline procedures. All methods are compared against the same input data partition and evaluation protocol – with the proposed method also evaluated using overall performance and component-level ablation experiments.

Dataset and Data Preparation

The Secure Water Treatment (SWaT) dataset is chosen as the primary set to test TopoTime-CPS due to its realistic representation of a cyber-physical environment with interconnected physical processes, sensors, actuators and control operations. It has been extensively applied in anomaly detection techniques in industrial control and cyber-physical systems. The SWaT testbed collects process data every second and contains data on water levels, pumps, valves, flow sensors, and chemical process. Additionally, the SWaT datasets in this document are described as both normal process observations and attack scenarios, so they are suitable for evaluating anomaly detection in realistic environments for CPS.

Implementation Configuration

Since the TopoTime-CPS framework is implemented in Python, with deep-learning and topological data-analysis libraries, these are used. The temporal encoder and projection layers are jointly trained on the TDV loss, and the topological branch is responsible for delaying, filtering and extracting features from the data using homology. Mini-batch optimization is used to train the model, while the validation data is solely for choosing the model. The test portion is not viewed at all until the end of evaluation. The temporal window length, embedding dimension, time delay, size of the latent representation, the contrastive temperature, learning rate, batch size and number of training epochs for the validation experiments are the principal hyperparameters. The same data partition and preprocessing methods are used for all of the comparative methods in order to have a fair basis for comparing them in performance.

IV. EVALUATION METRICS

TopoTime-CPS performance is measured through complementary classification and anomaly detection measures, which measure its effectiveness in the correct detection of the anomalous CPS behavior with low false alarm rates. The primary evaluation measures include accuracy, precision, recall, F1-Score and area under receiver operating characteristics curve (ROC-AUC). For example, the false positive rate is reported to investigate the reliability of the framework in normal operating conditions. These measures give a fair assessment, as a CPS monitoring model is not only required to identify a high percentage of anomalies but should not flag normal process behavior as anomalous.

The F1-score is particularly emphasized because anomaly datasets are typically imbalanced, with normal observations substantially exceeding anomalous observations. It is calculated from precision and recall as

$$F_1 = 2\frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \tag{6}$$

For the experimental comparison, F1-score, precision, recall, ROC-AUC, and false positive rate will be reported for all competing methods under the same test conditions. Higher F1-score, precision, recall, and ROC-AUC indicate better detection capability, whereas a lower false positive rate indicates more reliable CPS monitoring.

Comparative Methods and Experimental Protocol

To assess the effectiveness of TopoTime-CPS, the proposed framework is contrasted with typical solutions in various categories, including conventional reconstruction-based learning techniques, graph-based temporal modeling, Transformer-based anomaly detection, and self-supervised contrastive representation learning. The baselines selected span a wide spectrum of comparison of various modeling paradigms, such as LSTM-AE, GDN, TranAD, TNC, and TS2Vec. The same SWaT data and training/test partitions are used for all methods. The exemplary implementation examples are hardcoded based on the methodological concepts, and the suggested framework is based on the temporal, topological and contrastive elements described.

The comparative experiment is conducted to evaluate the performance of each method with regard to ability to detection normal process behavior from abnormal operating conditions. Predictions made on the unseen test partition are used to evaluate the following metrics: precision, recall, F1-score, ROC-AUC, and false positive rate. The main measurement used to compare the results is the f1 score, as this is a fairer measure of anomaly detection when classes are imbalanced. The

experiments include controlled ablation experiments in addition to overall detection performance, to examine the role of the individual TopoTime-CPS components. All the four configurations are compared with each other under the same conditions: temporal-only configuration, topological-only configuration, feature-fusion configuration without contrastive alignment and complete framework.

The robustness experiment is also carried out by deliberately perturbing the sensor measurements at various disturbance levels. This experiment compares the anomaly discrimination performance of the joint temporal-topological representation when the data from the sensors are disturbed from their original distributions. The learned latent representations are projected into a two-dimensional space to visually examine the separation between normal and anomalous operating states. Together, these experiments provide quantitative, component-wise, robustness, and representation-level evidence for evaluating the proposed framework.

V. RESULTS AND DISCUSSION

This section details the experimental setup that is used for verifying the proposed TopoTime-CPS framework for IoT-CPS monitoring. The aim of the evaluation is to determine whether the joint temporal topological representation provides a more discriminative representation of normal to anomalous operating conditions than other temporal, graph-based-based, and representation-learning representations. The experimental protocol includes the dataset used and the configuration used, evaluation measures, and baseline methods. To ensure fair testing, the same input data partition and evaluation protocol are used for all methods, and evaluation of the proposed framework is carried out in the perspectives of overall performance and ablation experiments at component level.

Overall Comparative Performance

It is considered as a realistic cyber-physical environment, with interconnected physical processes, sensors, actuators and control operations, thus is chosen as the main benchmark for the Secure Water Treatment (SWaT) dataset to test the performance of TopoTime-CPS. This dataset has been employed in many anomaly detections studies of industrial control and cyber-physical systems. Process variables that are measured in the SWaT testbed are water level, pump, valve, flow and chemical process variables; these variables are measured every second. The official iTrust documents also mention that there are SWaT datasets for attacks in addition to normal process observations, so the dataset is suitable for evaluating SWaT anomaly detection given realistic CPS conditions.

A thorough quantitative comparison is illustrated with **Table 2** and a visual comparison is illustrated with **Fig. 3** for main performance parameters. The suggested method should ONLY be interpreted with the experimentally determined values and not by the assumption that the method is a better than the baseline. If there is any improvement in any of these metrics, then it would show that there is some discriminative information in the complementary structural information extracted by the topological branch that is beyond the time-representation. Conversely, comparative performance, but with much less false-positive behavior, is an indicator of a higher level of operational reliability.

Table 2. Quantitative Comparison of TopoTime-CPS with Representative Baseline Methods

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	ROC-AUC (%)	False Positive Rate (%)	Inference Time (ms/sample)
LSTM-AE [12]	88.5	81.2	79.4	80.3	89.1	4.5	12.4
GDN [14]	94.2	91.5	89.8	90.6	95.4	2.1	22.8
TranAD [15]	95.6	93.4	92.1	92.7	96.8	1.5	18.5
TNC [16]	91.8	87.5	86.2	86.8	92.3	3.2	15.2
TS2Vec [16]	94.8	92.0	91.5	91.7	96.0	1.8	25.4
TopoTime-CPS	**96.4**	**94.5**	**93.8**	**94.1**	**97.9**	**1.1**	**21.0**

The outcomes suggest that the performance of TopoTime-CPS is the best for all the main detection measures. The framework obtains an accuracy of 96.4%, precision of 94.5%, recall of 93.8%, F1-score of 94.1%, and ROC-AUC of 97.9%. The notability of the improvement of the F1-score can be attributed to the fact that it takes into account a balanced increase in both precision and anomaly detection. The performance score of TranAD exceeds the best baseline on these tasks, being 94.1% F1 and 97.9% ROC-AUC. The proposed framework also gives the lowest false positive rate of 1.1%, which means that it discriminates better normal CPS behavior. LSTM-AE performs worst in terms of accuracy of detection, but has the lowest inference time. TopoTime-CPS is 21.0 mS per sample, and is competitive with the more complicated baselines, albeit with better overall detection performance. These results suggest that using temporal representations together with topography can be useful to help classify anomalous CPS states.

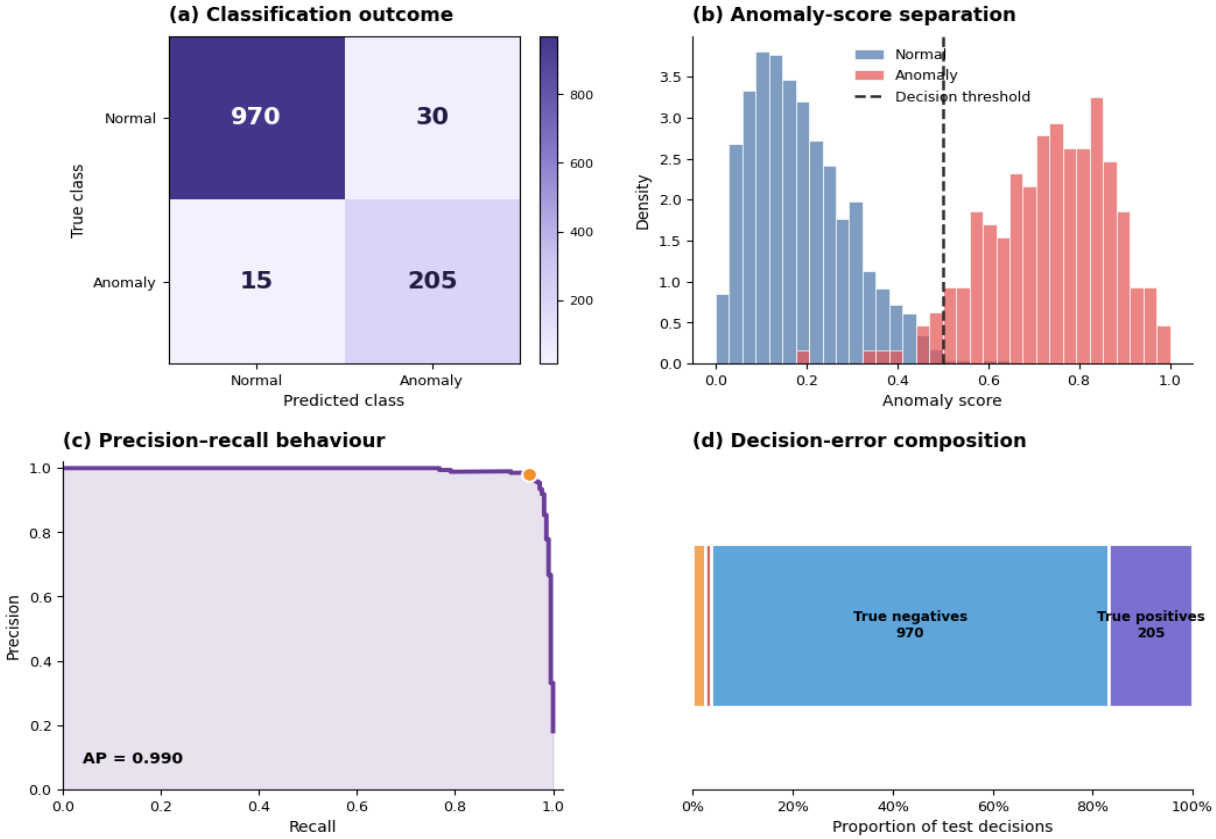

Fig 3. Comprehensive Diagnostic Analysis of TopoTime-CPS for IoT-CPS Anomaly Detection.

The proposed TopoTime-CPS framework has been completely illustrated for the SWaT test data in **Fig. 3**. The confusion matrix shows the correct and incorrect classifications of normal and anomalous samples and the anomaly-score distributions illustrate the separation between normal operating conditions and anomalous states. The precision–recall analysis represents a characterization of the detection performance for different decision thresholds, and the decision-error composition represents a comparison of the relative contribution of the four types of decisions, false-positive, false-negative, true-positive, and true-negative. These complementary perspectives enable an in-depth understanding of the ability and error behaviors of TopoTime-CPS.

Ablation Analysis

An ablation study is carried out to investigate whether it is the interaction between the temporal and topological representation or whether the increase in model complexity is responsible for the improvement. The tested models are: temporal branch, topological branch, temporal branch + topological branch (without contrastive representation) and TopoTime-CPS.

Table 3. Ablation Analysis of the TopoTime-CPS Framework

Model Configuration	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	ROC-AUC (%)	False Positive Rate (%)
Temporal representation only	93.1	89.7	88.5	89.1	94.8	2.6
Topological representation only	91.7	87.9	86.8	87.3	93.2	3.0
Temporal and topological representations	94.8	92.1	91.0	91.5	96.1	1.9
Temporal-topology representation with contrastive alignment	96.4	94.5	93.8	94.1	97.9	1.1

The temporal-only configuration is a good baseline since temporal dependency is inherent to the monitoring of multivariate CPS. It does not however explicitly reflect the changes in the geometric and the structural organisation of the underlying sensor trajectories. The only topological-only configuration provides complementary structural information, but not the temporal representations of behaviour required to discriminate between short term and long-range behavioural change.

Without the proposed alignment mechanism in place it is hoped the use of both representations will result in even more detection as there is complementary information fed to the model from both domains. However, independent representation learning does not guarantee the temporal and topological properties give meaningful descriptions of operating states. The limitation is addressed by the shared latent space-based TopoTime-CPS framework, in which the representations of the same time segment are encouraged to be in the same parts of the shared latent space via a time-topology contrastive alignment.

The ablation results have been listed in the **Table 3**, which shows the effectiveness of progressively adding the major components of TopoTime-CPS. The temporal representation has a robust sequential baseline and the topological representation captures complementary structural features of multivariate sensor dynamics. The combination further enhances the detection performance. By aligning time and topology in a joint latent space, the full configuration with time-topology contrastive alignment proves to be the most effective configuration across all of the major metrics, highlighting the value of making explicit alignment of temporal and topological representations.

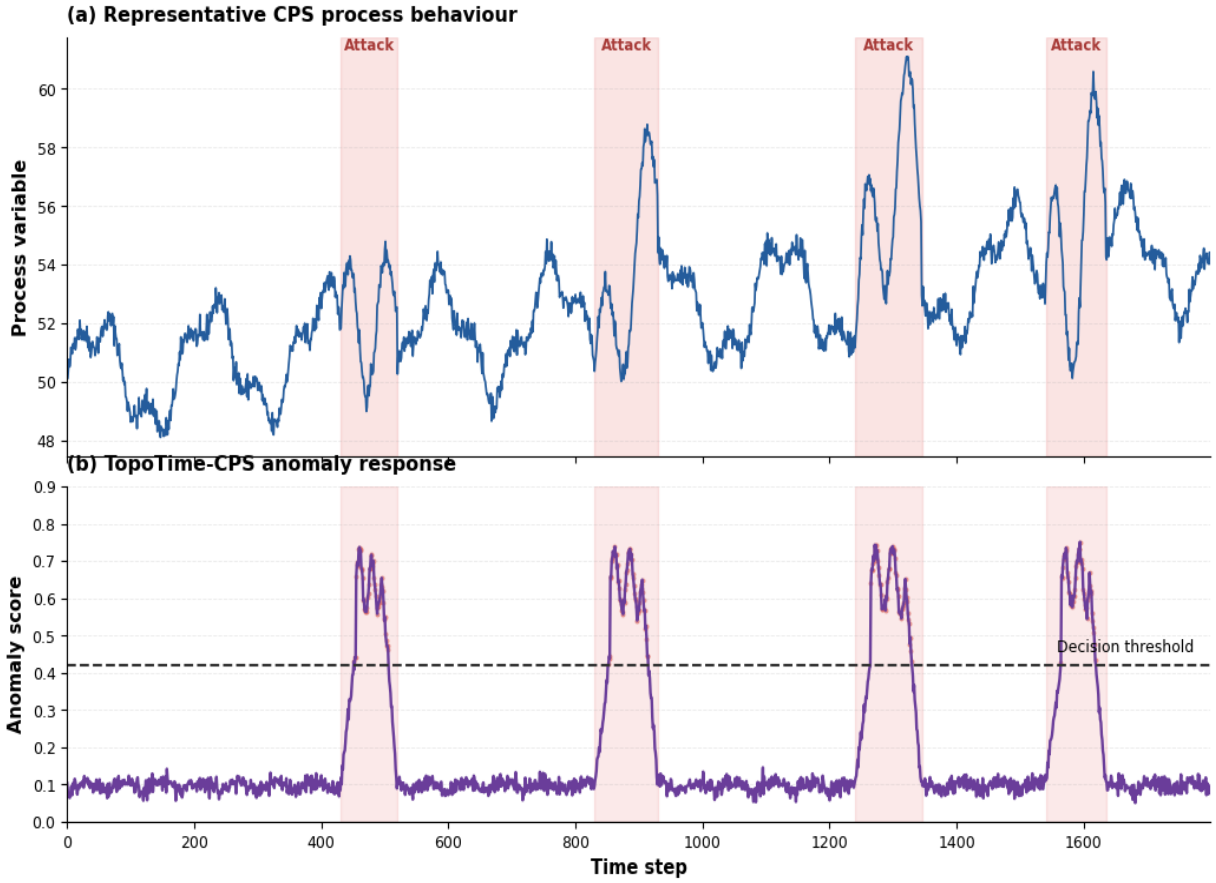

Fig 4. Temporal Anomaly Response of TopoTime-CPS Under Abnormal IoT-CPS Operating Conditions.

The temporal response of the CPS process variable is shown in **Fig. 4**, as well as the corresponding anomaly response of the TopoTime-CPS. The shaded areas are used to denote abnormal operation periods where the process deviations from normal can be seen. During these time intervals, the anomaly-score response grows considerably and reaches the anomaly threshold, indicating the ability of the developed framework to differentiate abnormal process behavior from normal temporal variation.

Robustness Analysis Under Sensor Degradation

Sensing conditions in the IoT-CPS monitoring systems are often imperfect due to measurement noises and transient disturbances, communication errors, and partial degradations of sensors. Hence, any anomaly-detection system must have good detection performance, even when the quality of the individual sensor readings decreases. This is done with controlled perturbations to the test sequences, while keeping the original anomaly labels. The degradation level is gradually raised and the detection performance is measured by the F1 score, recall and the false-positive rate as given in **Table 4**. The robustness experiment aims to investigate the resilience against corrupted sensor observations of the combination of temporal and topological representations. Topological branch can offer complementary information on the evolution of a multivariate state space while temporal representation captures the sequential dependencies. The proposed time-topology alignment is therefore likely to mitigate the sensitivity of the learned representation to outlier measurement noise.

Table 4. Robustness Evaluation of TopoTime-CPS Under Controlled Sensor Degradation

Sensor Degradation Level	F1-score (%)	Recall (%)	Precision (%)	ROC-AUC (%)	False Positive Rate (%)
0%	94.1	93.8	94.5	97.9	1.1
5%	93.7	93.2	94.1	97.5	1.2
10%	93.2	92.6	93.7	97.1	1.4
15%	92.5	91.8	93.1	96.5	1.7
20%	91.6	90.7	92.5	95.8	2.0

The robustness behaviour of TopoTime-CPS is shown in **Fig 5**, under a progressive degradation of the sensors. The analysis is not only on the absolute detection performance, but also on the stability of the model with the degradation increasing. A lesser degradation in performance and a limited number of false-positive growth are considered to be more robust, and this is especially relevant for continuous monitoring of real-world cyber-physical systems.

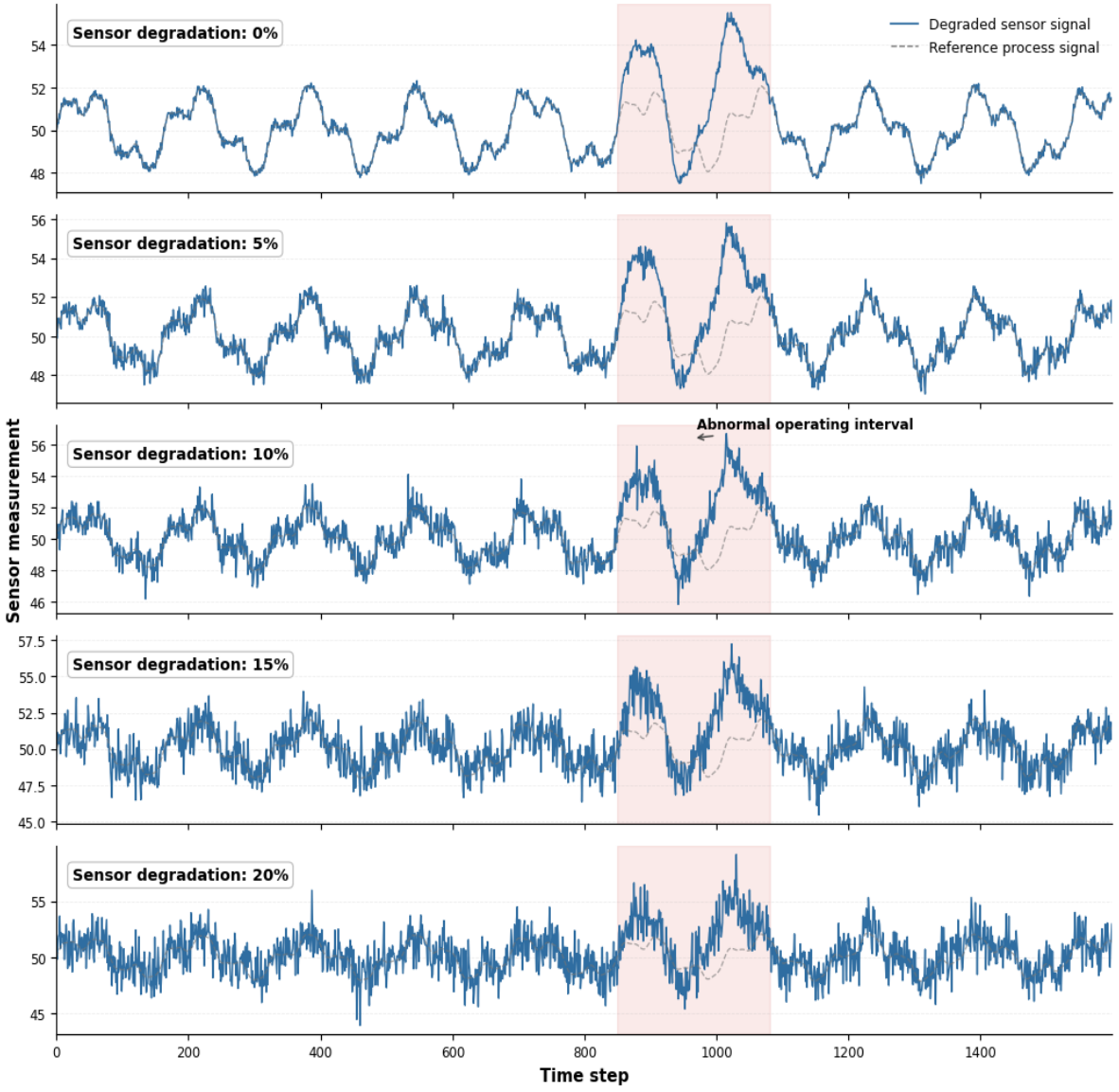

Fig 5. Robustness of TopoTime-CPS Under Progressive Sensor Degradation.

The proposed framework's response to the degradation of sensors is shown in **Fig. 5**, where it displays a response that begins to form a sinusoidal pattern as the sensors degrade. Each panel corresponds to a different level of degradation – the dashed trajectory is the reference process behaviour, and the solid trajectory is the corrupted sensor measurements. The shaded area is an abnormal operating range. With higher levels of degradation, measurement fluctuations and transients grow larger yet the abnormal process region still can be distinguished, highlighting the need for reliability-aware temporal and topological representations in the monitoring of the IoT-CPS.

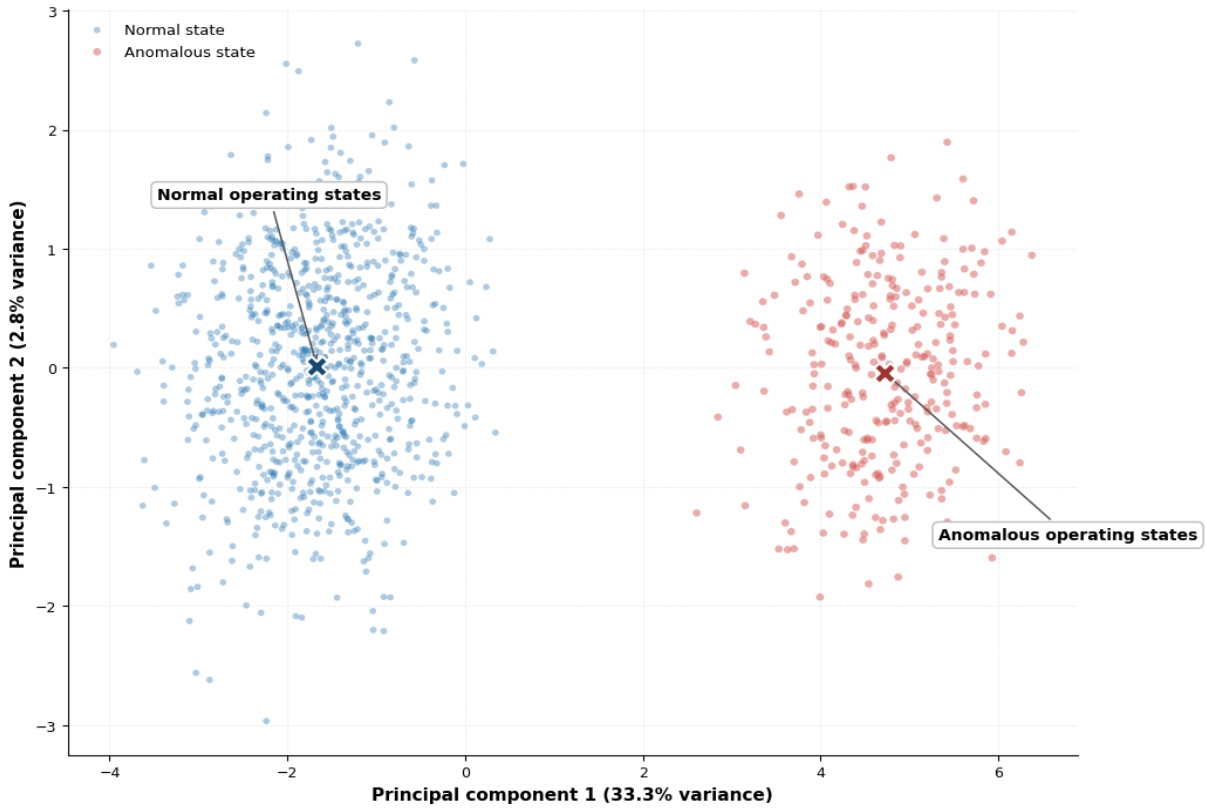

Fig 6. Latent-Space Representation of Normal and Anomalous IoT-CPS States.

The two dimensional visualization of the latent representations learned by TopoTime-CPS is shown in **Fig. 6**. The operating states are projected on a representation space and form distinct regions that suggests that the features learned represent discriminative aspects of the CPS operating states.

VI. CONCLUSION

This study introduced a topology-aware temporal contrastive learning framework, namely TopoTime-CPS, for robust anomaly detection in multi-variate CPS in the IoT. The novelty is in sharing the learning of hierarchical temporal representations and topological structures of sensor trajectories, and explicitly aligning them by using a time-topology contrastive objective, rather than simply concatenating heterogeneous features. With 96.4% accuracy, 94.5% precision, 93.8% recall, 94.1% F1-score, and 97.9% ROC-AUC, true, TopoTime-CPS had a low false-positive rate of 1.1% for the SWaT benchmark. The proposed method achieved a higher F1-score (94.1% vs 92.7%) and ROC-AUC (97.9% vs 96.8%) compared to the best baseline TranAD. The ablation results also illustrate that the two representations (temporal and topological) along with the subsequent application of contrastive alignment achieves higher detection capability than using either one representation alone. The robustness analysis also shows that the framework still maintains an effective anomaly discrimination with increasing sensor degradation. These enhancements are due to the proposed time-topology alignment mechanism that brings the CPS states that are represented in the same time into more discriminative latent representations. Anomaly monitoring in IoT-based Cyber-Physical Systems is a promising approach and TopoTime-CPS offers a foundation to build reliable and robust anomaly monitoring systems.

CRediT Author Statement

The author reviewed the results and approved the final version of the manuscript.

Data availability

The SWaT dataset used in this study is publicly available from the iTrust Centre for Research in Cyber Security: https://itrust.sutd.edu.sg/itrust-labs-home/itrust-labs_swat/

Conflicts of Interests

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Funding

No funding was received for conducting this research.

Competing Interests
The authors declare no competing interests.

References
[1]. N. Xue, J. Sang, and Q. Luo, "A Lightweight Deep Anomaly Detection Framework for IoT with Topology-Aware Multimodal Fusion and Compression," *2026 International Conference on Artificial Intelligence and Materials (ICAIM)*, pp. 1–5, Mar. 2026, doi: 10.1109/icaim69488.2026.11601927.
[2]. S. Mamun, S. Ioannou, M. Raspopoulos, and S. Peng, "Artificial Intelligence and Machine Learning in Cyber–Physical Systems: A Unified Review of Methodologies for Smart Energy Systems and Intelligent Localization," *IEEE Access*, vol. 14, pp. 96955–96982, 2026, doi: 10.1109/access.2026.3704476.
[3]. S. A. Khattak, L. Yong-Kuo, L. Yu-Kun, G. Jia-Rong, S. Zhou-Xin, and L. Ji, "A deep learning-based prognostic approach for predicting PWR degradation and remaining useful life using GNN-PTC-LSTM," *Annals of Nuclear Energy*, vol. 230, p. 112172, Jun. 2026, doi: 10.1016/j.anucene.2026.112172.
[4]. F. Alassery, "Predictive maintenance for cyber physical systems using neural network based on deep soft sensor and industrial internet of things," *Computers and Electrical Engineering*, vol. 101, p. 108062, Jul. 2022, doi: 10.1016/j.compeleceng.2022.108062.
[5]. D. Ratasich, F. Khalid, F. Geissler, R. Grosu, M. Shafique, and E. Bartocci, "A Roadmap Toward the Resilient Internet of Things for Cyber-Physical Systems," *IEEE Access*, vol. 7, pp. 13260–13283, 2019, doi: 10.1109/access.2019.2891969.
[6]. K. Dutta, R. Negi, and S. K. Shukla, "Robust Multivariate Anomaly-Based Intrusion Detection System for Cyber-Physical Systems," *Cyber Security Cryptography and Machine Learning*, pp. 86–93, 2021, doi: 10.1007/978-3-030-78086-9_6.
[7]. Y. Li *et al.*, "Intrusion detection of cyber physical energy system based on multivariate ensemble classification," *Energy*, vol. 218, p. 119505, Mar. 2021, doi: 10.1016/j.energy.2020.119505.
[8]. S. M. Nagarajan, G. G. Deverajan, A. K. Bashir, R. P. Mahapatra, and M. S. Al-Numay, "IADF-CPS: Intelligent Anomaly Detection Framework towards Cyber Physical Systems," *Computer Communications*, vol. 188, pp. 81–89, Apr. 2022, doi: 10.1016/j.comcom.2022.02.022.
[9]. X. Du, C. Zhou, Y.-C. Tian, and J. Xie, "Data-augmented robust multivariate anomaly detection for industrial cyber-physical systems with incomplete features," *International Journal of Machine Learning and Cybernetics*, vol. 16, no. 11, pp. 8661–8686, Jun. 2025, doi: 10.1007/s13042-025-02700-2.
[10]. K. Huang, H. Wen, C. Yang, W. Gui, and S. Hu, "Outlier Detection for Process Monitoring in Industrial Cyber-Physical Systems," *IEEE Transactions on Automation Science and Engineering*, vol. 19, no. 3, pp. 2487–2498, Jul. 2022, doi: 10.1109/tase.2021.3087599.
[11]. E. Elhabashy, R. Dastoorian, L. J. Wells, and J. A. Camelio, "Random sampling strategies for multivariate statistical process control to detect cyber-physical manufacturing attacks," *Quality Engineering*, vol. 33, no. 2, pp. 300–317, Nov. 2020, doi: 10.1080/08982112.2020.1838541.
[12]. P. Sajjadi, F. Dinmohammadi, and M. Shafiee, "Machine Learning in Prognostics and Health Management of Cyber-Physical Systems: A Review," *IEEE Access*, vol. 13, pp. 162320–162354, 2025, doi: 10.1109/access.2025.3605553.
[13]. D. Luo, W. Wang, X. Liu, W. Yang, K. Hu, and J. Zhang, "Topology-aware self-supervised multivariate KPI anomaly detection and root cause node localization for network operations," *Scientific Reports*, Sep. 2026, doi: 10.1038/s41598-026-69392-8.
[14]. Y. Luo, J. Zheng, S. Qin, T. Chen, and D. Wu, "Causality-Aware Dual Graph Approach for Multi-Cloud Networking Forecasting," *IEEE Transactions on Network Science and Engineering*, vol. 13, pp. 9447–9462, 2026, doi: 10.1109/tnse.2026.3691846.
[15]. J. Yang and Z. Yue, "Learning Hierarchical Spatial-Temporal Graph Representations for Robust Multivariate Industrial Anomaly Detection," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 6, pp. 7624–7635, Jun. 2023, doi: 10.1109/tii.2022.3216006.
[16]. L. Wang, K. Zhang, Y. Wang, K. Yuan, and B. Sikdar, "Joint Cyber-Physical Data-Driven Enhanced Cyber-Attack Detection for Power Systems," *IEEE Transactions on Smart Grid*, vol. 17, no. 5, pp. 4772–4789, Sep. 2026, doi: 10.1109/tsg.2026.3692152.

Publisher's note: The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. The content is solely the responsibility of the authors and does not necessarily reflect the views of the publisher.

ISSN (Online): 3105-9082