CONTRASTIVE LEARNING OF DATA-QUALITY-INVARIANT REPRESENTATIONS FOR TIME SERIES CLASSIFICATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Data quality issues are pervasive in time series collected from real-world sensing systems. Such issues can induce distribution shift and degrade downstream task performance. In this paper, we propose a self-supervised contrastive learning framework to learn data-quality-invariant representations for time series classification. First, we propose anomaly injection to generate negative samples for contrast with normal samples in the latent space, thereby mitigating data-quality-induced distribution shift. Second, we use autoencoder-based reconstruction as the pretext task to learn discriminative representations. Lastly, we freeze the pre-trained backbone and employ a prototypical network for few-shot time series classification using a small number of labeled samples. Extensive experiments on real-world datasets demonstrate that our proposed model consistently outperforms state-of-the-art self-supervised contrastive learning frameworks, improving accuracy by an average of 4.09 percentage points across the three datasets compared with the strongest baseline on each dataset. An ablation study also shows that anomaly injection is effective at improving classification performance compared with current data augmentation techniques.

1 INTRODUCTION

Time series classification is a fundamental task in a wide variety of real-world applications, including industrial monitoring, healthcare, and asset condition assessment (Wang et al., 2024). Despite advances in deep learning for time series, distribution shifts arising from data quality issues pose a critical challenge to robust classification (Kulinski et al., 2020; Li et al., 2021; Liu et al., 2026). In practical sensing systems, multiple complex factors, such as sensor faults, hardware degradation, and temperature fluctuations, can lead to abnormal measurements. As a result, samples deviate from their original class distributions and potentially are misclassified into other classes.

Self-supervised contrastive representation learning has emerged as an effective paradigm for learning informative temporal representations from unlabeled time series data. Existing methods typically construct positive pairs using data augmentations or temporal neighborhoods, while treating other instances or temporally distant samples as negatives (Tonekaboni et al., 2021; Zheng et al., 2024; Lee et al., 2024b). However, these pairing strategies are not explicitly designed to learn representations that are invariant to data-quality issues, because an augmented view may not be positive, whereas a temporally distant sample may still preserve similar underlying dynamics under data-quality perturbations.

In this paper, we address this problem from the perspective of data-quality-invariant representation learning. Our key idea is to explicitly introduce realistic data-quality perturbations in the latent space during self-supervised pre-training. Specifically, we propose a reconstruction-based contrastive learning framework, where normal temporal windows serve as anchors, neighboring windows provide semantically related positive samples, and anomaly-injected counterparts serve as neg-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

ative samples. The injected anomalies represent realistic abnormal data observed from the sensing system. By contrasting intrinsic temporal patterns with diverse data-quality-induced perturbations, the model preserves discriminative temporal information and disentangles it from quality-related variations. A second challenge in real-world time series classification is the limited availability of reliable labels, as annotation often requires considerable human effort and domain expertise. To reduce this dependence, we transfer the representations learned from unlabeled data to a few-shot classification setting. After self-supervised pre-training, the encoder is frozen and combined with a Prototypical Network, where a small support set is used to construct class prototypes and query samples are classified based on their proximity to these prototypes.

Our main contributions are summarized as follows:

- We formulate time series classification under data-quality-induced distribution shifts and introduce a data-quality-invariant self-supervised representation learning framework that explicitly models realistic anomalies during pre-training.
- We develop a reconstruction-based contrastive learning strategy that combines anomaly injection with temporal contrastive learning to preserve class-discriminative temporal information while separating it from quality-related variations.
- We integrate the learned representations with prototype-based few-shot classification, enabling effective downstream recognition with limited labeled data, and demonstrate its effectiveness on multiple real-world datasets.

2 Related Work

Self-supervised time series representation Self-supervised representation learning has achieved considerable success across a variety of domains, including computer vision (Chen et al., 2020; Milbich et al., 2020), natural language processing (Logeswaran & Lee, 2018), and speech recognition (Xu et al., 2021), thereby motivating its application to time-series analysis. Representative approaches such as TKAE (Bianchi et al., 2019) and TST (Zerveas et al., 2021) have been developed to learn informative temporal representations without extensive supervision. However, these methods may be limited in scalability when handling very long sequences or in capturing complex temporal dynamics. To address these issues, TNC (Tonekaboni et al., 2021) and T-Loss (Franceschi et al., 2019) exploit temporal neighborhoods, negative sampling, triplet-based objectives, and local smoothness to improve the scalability and effectiveness of multivariate time-series representation learning. Nevertheless, many existing methods either encode representations at specific semantic levels or rely on strong assumptions of invariance under predefined transformations, which may limit their generalizability. More recently, methods such as RoSAS (Xu et al., 2023) have explored triplet-based representation learning for outlier detection beyond conventional time-series settings.

Contrastive representation learning Contrastive learning has become a widely used paradigm for representation learning by promoting consistency among related samples and separation among unrelated ones. Classical approaches such as InfoNCE (Oord et al., 2018) and SimCLR (Chen et al., 2020) demonstrated the effectiveness of this principle through pairwise contrastive objectives. In time-series analysis, the same idea has been extended to anomaly detection by exploiting temporal structure. TS2Vec (Yue et al., 2022) learns hierarchical representations across multiple temporal scales, while DCdetector (Yang et al., 2023) employs asymmetric contrastive modeling over dual-attention representations to capture anomalous temporal dependencies.

Reconstruction-based representation learning Reconstruction provides a self-supervised objective for learning temporal representations by recovering original or masked observations from partially corrupted inputs. Zerveas et al. (2021) introduced a Transformer-based framework for multivariate time series, where segments of the input are masked and the model is trained to recover the missing values, enabling transferable representations for downstream classification and regression. PatchTST (Nie et al., 2022) further adopts patch-level masked pre-training and demonstrates strong transfer and fine-tuning performance for time-series representation learning. SimMTM (Dong et al., 2023) revisits masked time-series modeling by noting that random point masking may destroy important temporal variations, and reconstructs masked observations through the aggregation of

complementary masked series. More recently, PITS (Lee et al., 2024a) argues that explicitly modeling dependencies across masked patches is not always necessary and instead learns representations through independent patch reconstruction complemented by contrastive learning. While these approaches demonstrate the effectiveness of reconstruction as a self-supervised learning pretext task, reconstruction objectives primarily focus on recovering temporal content and do not explicitly structure the latent space with respect to data-quality perturbations. In contrast, our framework uses reconstruction to preserve intrinsic temporal information while introducing anomaly-aware contrastive constraints to explicitly model data-quality variations.

3 METHODOLOGY

3.1 OVERVIEW

Given a dataset of N time-series instances $\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$, each sequence $\mathbf{x}_i \in \mathbb{R}^{T \times D}$, with length T and dimensionality D, is partitioned into K overlapping windows $\mathbf{x}_{i,j}$ with an overlap ratio p. The window length is given by $L = \frac{T}{1+(K-1)(1-p)}$ corresponding to a stride of $L(1-p)$. Our model learns a temporal encoder that maps the input time series into data-quality-invariant latent representations for robust classification.

The overall framework of our two-stage model is illustrated in Figure 1. Initially, we employ anomaly injection to extract data-quality-invariant representations with reconstruction-based contrastive learning. These injected anomalies are derived from observations of faulty sensor data and represent universal data quality issues in real-world scenarios. The design of anomaly injection aims to mitigate data-quality-induced distribution shifts by contrasting original windows with their anomaly-injected counterparts. The few-shot classifier then assigns each window to its category based on its proximity to class-centric representations in the latent space, leveraging the learned data-quality invariance to preserve class consistency under data-quality issues.

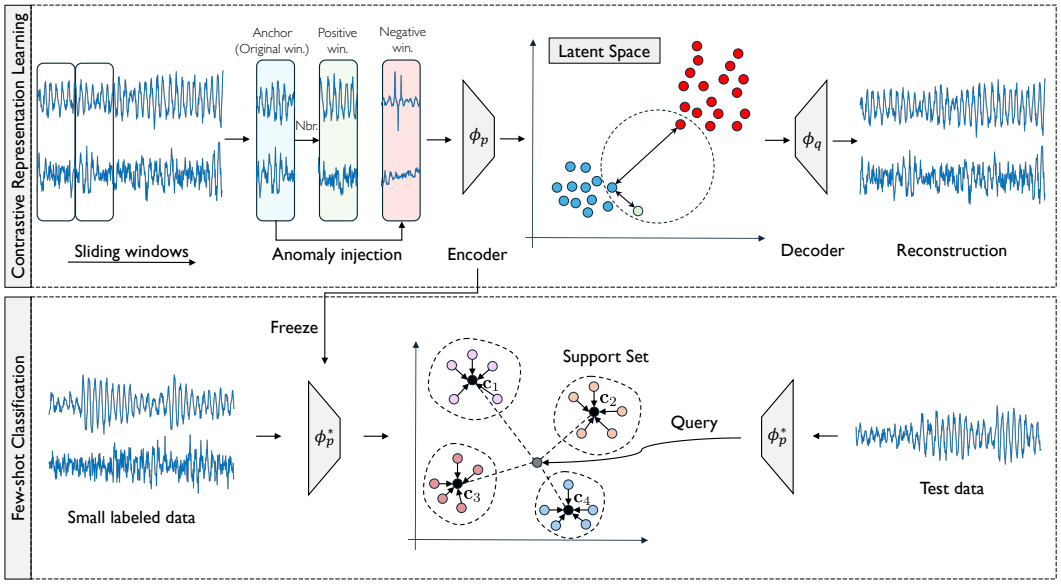


Figure 1: Overview of the framework. During pre-training, each window serves as an anchor, with a neighboring window as the positive and its anomaly-injected counterparts as negatives. The encoder ϕ_p maps these windows into a latent space by anchor and positive against negative, while the decoder ϕ_q reconstructs the original signal to preserve temporal information. For downstream classification, the pre-trained encoder ϕ_p^* is frozen, class prototypes are computed from the support set, and query samples are classified according to their proximity to the prototypes in the latent space.

3.2 ANOMALY INJECTION

We implement anomaly injection as a data augmentation technique to create negative samples for contrastive learning (Darban et al., 2025). Each window is processed by randomly injecting abnormal data from a specific anomaly type. In existing time series contrastive learning methods, positive samples are augmented windows, while temporally distant windows are defined as negative samples. We argue that these assumptions are limited in the context of learning data-quality-invariant representations, as time series augmentation may transform them into negative samples, and a temporally distant window may preserve consistent dynamics of a positive sample. In contrast, anomaly injection enables our model to mitigate the effects of data anomalies in time series classification.

Our anomaly injection method introduces pointwise anomalies into normal data at randomly selected time steps within a given window. We employ five distinct types of anomalies, including `Outlier`, `Drift`, `Bias`, `Minor`, and `Noise`. These anomalies are often observed in practical sensor data (Pan et al., 2026) and represent typical manifestations of data-quality issues. We aim to learn semantically structured and data-quality-invariant representations by explicitly modeling diverse anomaly effects during training. We elaborate further on the details of anomaly injection in Appendix A.

3.3 RECONSTRUCTION-BASED CONTRASTIVE REPRESENTATION LEARNING

In real-world applications, data-quality issues can induce distribution shifts in sensor data, causing samples to deviate from their original class distributions and potentially be misclassified by purely reconstruction-based methods. To address this issue, we propose a reconstruction-based contrastive representation learning model that extracts data-quality-invariant representations while preserving class-discriminative information for robust time series classification.

We implement an encoder-decoder architecture as the reconstruction backbone, which is trained exclusively on normal data. Specifically, for each window j of sample i, the encoder ϕ_p compresses the input time series $\mathbf{x}_{i,j}$ into a latent space as representations $\mathbf{z}$, followed by a decoder ϕ_q that reconstructs the original time series window with $\mathbf{z}$. Both ϕ_p and ϕ_q can be instantiated with different backbones for extracting temporal representations, such as InceptionTime (Ismail Fawaz et al., 2020), Transformer (Vaswani et al., 2017), and LSTM (Hochreiter & Schmidhuber, 1997). The reconstruction backbone is trained by minimizing discrepancies between the original and reconstructed time series:

$$\mathcal{L}_{\text{Recon}} = \frac{1}{NK} \sum_{i=1}^{N} \sum_{j=1}^{K} \|\phi_q(\phi_p(\mathbf{x}_{i,j})) - \mathbf{x}_{i,j}\|_2^2. \tag{1}$$

To achieve data-quality-invariant representations, we impose self-supervised contrastive constraints on $\mathbf{z}$ in the latent space. The original window $\mathbf{x}_{i,j}$ is treated as the anchor $\mathbf{a}_{i,j}$. A positive window $\mathbf{p}_{i,j}$ is selected from the anchor's neighboring windows other than itself, i.e., $\mathbf{p}_{i,j} = \mathbf{x}_{i,k}$, with $k \sim \mathcal{U}(\{1, \ldots, K\} \setminus \{j\})$, while a set of negative windows $\{\mathbf{n}_{i,k}\}_{k=1}^{K}$ is generated by injecting anomalies into the anchor window. We use temporal instance classification as the pretext task for self-supervised contrastive learning, encouraging representations of anchors and their positive windows to be similar while contrasting them against all other negative windows in the sequence. This ensures that samples affected by data-quality-induced distribution shifts yield representations similar to those of their original classes. The representations of the anchor, positive, and negative windows are learned from the encoder ϕ_p:

$$\mathbf{z}_{i,j}^{a} = \phi_p(\mathbf{a}_{i,j}), \quad \mathbf{z}_{i,j}^{+} = \phi_p(\mathbf{p}_{i,j}), \quad \mathbf{z}_{i,k}^{-} = \phi_p(\mathbf{n}_{i,k}), k \in \{1, \ldots, K\}. \tag{2}$$

The contrastive learning objective is then defined using the InfoNCE loss (Oord et al., 2018):

$$\mathcal{L}_{\text{InfoNCE}} = \frac{1}{NK} \sum_{i=1}^{N} \sum_{j=1}^{K} -\log \frac{\exp\left(\mathbf{z}_{i,j}^{a} \cdot \mathbf{z}_{i,j}^{+}/\tau\right)}{\exp\left(\mathbf{z}_{i,j}^{a} \cdot \mathbf{z}_{i,j}^{+}/\tau\right) + \sum_{k=1}^{K} \exp\left(\mathbf{z}_{i,j}^{a} \cdot \mathbf{z}_{i,k}^{-}/\tau\right)}, \tag{3}$$

where τ is a temperature hyperparameter. This contrastive loss is the cross-entropy loss of a (K+1)-way classifier that intends to identify the anchor as its positive window.

A joint loss, $\mathcal{L} = \mathcal{L}_{\text{Recon}} + \lambda \mathcal{L}_{\text{InfoNCE}}$, is adopted as the training objective for reconstruction-based contrastive representation learning, where λ is a weighting coefficient that balances the contributions of the reconstruction and contrastive learning objectives.

3.4 Few-shot Classification

The primary objective of contrastive learning with anomaly injection is to acquire data-quality-invariant representations from unlabeled data, thereby reducing the reliance on large-scale manual annotation and domain expertise. We freeze the pre-trained backbone and employ a Prototypical Network (Snell et al., 2017) for few-shot time series classification using a small labeled subset. We first organize the samples into a support set $\mathcal{S}$ and a query set $\mathcal{Q}$, where $\mathcal{S} \cap \mathcal{Q} = \varnothing$, followed by a frozen encoder ϕ_p^* to map each support sample into the data-quality-invariant representations. For each class c, the prototype is computed by averaging these representations:

$$\mathbf{c}_c = \frac{1}{|\mathcal{S}_c|} \sum_{(\mathbf{x},y)\in\mathcal{S}_c} \left(\phi_p^*(\mathbf{x})\right), \tag{4}$$

where $\mathcal{S}_c$ denotes the subset of support samples for class c. Given a query sample $\mathbf{x}_q$, its representation $\phi_p^*(\mathbf{x}_q)$ is compared with each class prototype using a distance function $d(\cdot,\cdot)$. The query sample $\mathbf{x}_q$ is predicted as the class with the highest probability over C classes:

$$\hat{y}_q = \underset{c\in\{1,\ldots,C\}}{\arg\max}\ p(y=c|\mathbf{x}_q) = \underset{c\in\{1,\ldots,C\}}{\arg\max} \frac{\exp -d\left(\phi_p^*(\mathbf{x}_q), \mathbf{c}_c\right)}{\sum_{i=1}^{C} \exp -d\left(\phi_p^*(\mathbf{x}_q), \mathbf{c}_i\right)}. \tag{5}$$

4 Experiments

4.1 Experimental setup

Datasets Three structural vibration datasets of Z24 Bridge Dataset, IASC-ASCE Dataset, and the four-story steel frame experimental dataset are employed. These datasets cover different types of structural condition variations, including bridge pier settlement, structural component damage, and loosening of structural connection bolts. In addition, the effects of sensor anomalies on the acquired monitoring signals are taken into consideration. To ensure a consistent experimental setup across the three datasets, fixed-length segments are extracted from the continuous vibration signals under each structural condition. Based on the resulting structural-state samples, two types of sensor conditions, namely normal and anomalous, are further constructed. The samples are ultimately categorized into four states, i.e., healthy-normal, healthy-anomalous, damaged-normal, and damaged-anomalous.

Baselines To evaluate the effectiveness of the proposed method for vibration-based structural damage detection, a comparative study was conducted with several representative self-supervised time-series representation learning methods. The selected baselines cover recent contrastive learning approaches that exploit different temporal characteristics, frequency information, positive-pair construction strategies, and representation learning mechanisms. Specifically, PLanTS (Wang et al., 2025), Di-COT (Shamba et al., 2026), MF-CLR (Duan et al., 2024), DBPM (Lan et al., 2024), RE-BAR (Xu et al., 2024), AutoTCL (Zheng et al., 2024), SoftCLT (Lee et al., 2024b), TS2Vec (Yue et al., 2022), CoST (Woo et al., 2022), and TimesURL (Liu & Chen, 2024) were selected for comparison.

Implementation Details Unless otherwise specified, the same hyperparameter settings were used across the three datasets. The batch size was set to 8, and the output representation dimension was fixed at 128. The network depth was set to 3, with a learning rate of 3×10^{-4} and a weight decay of 3×10^{-4}. The model was trained for 30 epochs with 3000 training iterations. The overlap probability was set to $p = 0.5$ for all three datasets. The temperature parameter was set to $\tau = 0.07$ for Steel Frame and $\tau = 0.05$ for IASC-ASCE and Z24. A random seed of 42 was used throughout the

experiments. For the few-shot evaluation, $K = 5$ samples from each class were randomly selected as the support set, while the remaining samples were used for testing.

4.2 RESULTS

To further evaluate the effectiveness of the learned representations for structural damage identification, we compare the proposed method with several representative self-supervised time-series representation learning methods on three structural vibration datasets. The comparison is conducted using Accuracy, Precision, Recall, and F1-score, with the results summarized in Table 1. Unlike conventional time-series representation learning methods that mainly rely on a single contrastive objective, the proposed framework integrates InceptionTime-based multi-scale temporal feature extraction, temporal contrastive learning, data-quality-invariant representation learning, and autoencoder-based reconstruction. This design allows the learned representations to capture structural vibration patterns at different temporal scales while retaining robustness to sensor disturbances and preserving informative signal characteristics.

Table 1: Classification performance (%) on three datasets.

Dataset	Method	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Steel Frame	AutoTCL	65.45	65.79	65.45	65.38
	CoST	25.45	30.18	25.45	26.93
	DBPM	60.45	61.37	60.45	60.17
	Di-COT	75.91	76.50	75.91	75.90
	MF-CLR	73.64	75.10	73.64	73.34
	PLanTS	77.27	78.11	77.27	77.17
	REBAR	59.55	59.88	59.55	57.14
	SoftCLT	72.27	72.34	72.27	72.27
	TimesURL	78.18	79.43	78.18	78.00
	TS2Vec	68.64	68.88	68.64	68.67
	Proposed Method	**84.55**	**85.95**	**84.55**	**84.35**
IASC-ASCE	AutoTCL	63.64	63.88	63.64	63.56
	CoST	47.73	48.20	47.73	47.75
	DBPM	55.45	55.89	55.45	55.36
	Di-COT	72.27	73.29	72.27	71.86
	MF-CLR	86.36	87.62	86.36	86.24
	PLanTS	57.73	58.33	57.73	56.42
	REBAR	63.18	63.69	63.18	62.84
	SoftCLT	61.82	65.92	61.82	59.76
	TimesURL	81.36	83.46	81.36	81.09
	TS2Vec	86.82	88.91	86.82	86.52
	Proposed Method	**90.45**	**90.90**	**90.45**	**90.43**
Z24	AutoTCL	65.00	65.27	65.00	64.22
	CoST	35.00	41.81	35.00	34.10
	DBPM	40.91	39.49	40.91	39.58
	Di-COT	66.82	67.17	66.82	66.79
	MF-CLR	59.55	60.27	59.55	58.86
	PLanTS	68.18	69.00	68.18	67.86
	REBAR	43.18	46.10	43.18	35.94
	SoftCLT	56.82	59.25	56.82	56.32
	TimesURL	50.91	50.80	50.91	49.77
	TS2Vec	66.36	66.94	66.36	66.03
	Proposed Method	**70.45**	**70.54**	**70.45**	**70.43**

On the Steel Frame dataset, the proposed method achieves the highest performance among all compared methods, with an Accuracy of 84.55%, Precision of 85.95%, Recall of 84.55%, and F1-score of 84.35%. Compared with the strongest baseline, TimesURL, the proposed method improves Accuracy and F1-score by 6.37 and 6.35 percentage points, respectively. The improvement suggests that multi-scale temporal feature extraction is beneficial for the experimental vibration signals, where structural responses may exhibit characteristic patterns over different temporal scales. The combi-

nation of temporal contrastive learning and reconstruction further constrains the representation to preserve both temporal consistency and informative vibration characteristics.

On the IASC-ASCE dataset, the proposed method achieves the best overall performance among all compared methods, obtaining 90.45% Accuracy, 90.90% Precision, 90.45% Recall, and 90.43% F1-score. Among the baseline methods, TS2Vec provides the strongest performance, with an Accuracy of 86.82%, Precision of 88.91%, Recall of 86.82%, and F1-score of 86.52%. Relative to TS2Vec, the proposed method improves Accuracy and F1-score by 3.63 and 3.91 percentage points, respectively. The substantial improvement on the IASC-ASCE dataset indicates that the proposed representation learning strategy is effective in handling vibration signals with an increased number of channels. In particular, the multi-scale convolutional backbone can integrate vibration information across different temporal patterns.

The Z24 dataset presents a different performance profile. Compared with the other two datasets, the absolute performance of most methods is lower, and the differences between several competing approaches are relatively small. Nevertheless, the proposed method achieves the highest results across all four metrics, with 70.45% Accuracy, 70.54% Precision, 70.45% Recall, and 70.43% F1-score. Among the baseline methods, PLanTS provides the strongest overall performance, achieving 68.18% Accuracy, 69.00% Precision, 68.18% Recall, and 67.86% F1-score. Consequently, the proposed method improves Accuracy and F1-score by 2.27 and 2.57 percentage points, respectively. Although the margin is smaller than that observed on the Steel Frame and IASC-ASCE datasets, the proposed method remains consistently superior under the more challenging Z24 condition. In particular, the close values of Accuracy and Recall, together with the F1-score of 70.43%, indicate that the improvement is relatively balanced rather than being driven by a single evaluation metric.

The results across the three datasets also reveal substantial differences among the baseline methods. For example, MF-CLR performs strongly on the IASC-ASCE dataset, reaching 86.36% Accuracy and 86.24% F1-score, but its performance decreases considerably on the Z24 dataset, where the corresponding values are 59.55% and 58.86%. TimesURL exhibits a similar dataset-dependent behavior, obtaining 78.18% Accuracy on Steel Frame and 81.36% on IASC-ASCE, but only 50.91% on Z24. In contrast, TS2Vec performs particularly well on IASC-ASCE, while PLanTS provides the strongest baseline results on Z24. These variations indicate that no single baseline representation exhibits the same level of effectiveness across all structural systems. The differences are likely related to the distinct temporal characteristics, measurement conditions, and structural responses represented in the three datasets.

The relatively consistent performance of the proposed method can be interpreted from the complementary nature of its learning objectives. Multi-scale temporal feature extraction enables the model to represent vibration patterns with different temporal extents, while temporal contrastive learning encourages semantically related observations to remain close in the learned feature space. The anomaly-aware component further introduces sensitivity to irregular observations, which is particularly relevant to vibration measurements that may contain disturbed or atypical responses. Meanwhile, autoencoder-based reconstruction imposes an information-preserving constraint on the learned representation rather than optimizing solely for discriminative separation. Rather than relying on any single objective, the proposed framework therefore learns the representation from several complementary perspectives.

An important observation from Table 1 is that the advantage of the proposed method is evident not only in Accuracy but also in F1-score. This is relevant for structural damage identification because a high overall classification accuracy alone may not fully reflect the ability to distinguish different structural conditions. The consistently higher Precision, Recall, and F1-score obtained by the proposed method indicate that its performance gain is accompanied by a more balanced classification behavior. The relatively small gap between Accuracy and F1-score on all three datasets further suggests that the observed improvement is not primarily caused by a favorable bias toward a particular class.

Overall, the results in Table 1 provide consistent empirical evidence that the proposed representation learning framework can produce discriminative features for structural vibration-based damage identification across different structural systems. The proposed method ranks highest across all four evaluation metrics on the Steel Frame, IASC-ASCE, and Z24 datasets, while the magnitude of improvement varies according to the dataset. The larger gains on Steel Frame and IASC-ASCE and the more moderate improvement on Z24 suggest that the effectiveness of the learned representation is influenced by the underlying signal characteristics, rather than being independent of the dataset. Nevertheless, the consistent improvement over different baseline methods indicates that combining multi-scale temporal modeling, contrastive learning, anomaly-aware representation learning, and autoencoder-based reconstruction provides a useful basis for extracting robust representations from structural

vibration signals. These learned representations can subsequently serve as an informative feature space for downstream structural damage identification tasks.

4.3 Ablation Study

To investigate the contribution of the key design choices in the proposed framework, an ablation study was conducted on the Steel Frame, IASC-ASCE, and Z24 datasets. Three variants were considered in addition to the full model. First, the autoencoder-based reconstruction module was removed while keeping all other components unchanged, denoted as w/o Autoencoder-based reconstruction. Second, the anomaly injection strategy was removed, denoted as w/o Anomaly Injection. Third, the original temporal feature extraction backbone was replaced with a Transformer-based backbone to examine the effectiveness of the adopted architecture for structural vibration representation learning. Accuracy and F1-score were used as the evaluation metrics, and the results are summarized in Table 2.

Table 2: Ablation study of the proposed method on three datasets.

Dataset	Model Variant	Accuracy (%)	F1-score (%)
Steel Frame	Full Model	**84.55**	**84.35**
	w/o Autoencoder-based reconstruction	83.67	83.29
	w/o Anomaly Injection	82.73	82.46
	Replace the backbone with Transformer	81.82	81.55
IASC-ASCE	Full Model	**90.45**	**90.43**
	w/o Autoencoder-based reconstruction	88.48	88.18
	w/o Anomaly Injection	87.73	86.49
	Replace the backbone with Transformer	83.64	83.23
Z24	Full Model	**70.45**	**70.43**
	w/o Autoencoder-based reconstruction	70.21	69.64
	w/o Anomaly Injection	69.55	68.92
	Replace the backbone with Transformer	69.09	68.91

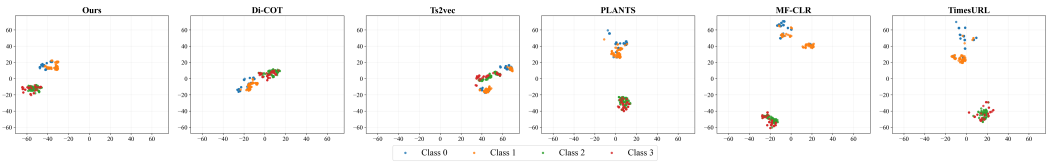
The full model consistently achieves the highest Accuracy and F1-score across all three datasets. On the Steel Frame dataset, the full model achieves an Accuracy of 84.55% and an F1-score of 84.35%. Removing the autoencoder-based reconstruction module results in moderate decreases of 0.88 and 1.06 percentage points in Accuracy and F1-score, respectively. A larger degradation is observed when anomaly injection is removed, with Accuracy and F1-score decreasing by 1.82 and 1.89 percentage points. Replacing the backbone with a Transformer leads to a larger decrease, with Accuracy and F1-score dropping to 81.82% and 81.55%. This indicates that both components contribute to the learned representation, while the adopted backbone has a relatively stronger effect on this dataset.

A similar trend is observed on the IASC-ASCE dataset. The full model achieves 90.45% Accuracy and 90.43% F1-score, whereas removing the autoencoder-based reconstruction module reduces the two metrics to 88.48% and 88.18%, corresponding to decreases of 1.97 and 2.25 percentage points. Without anomaly injection, the Accuracy and F1-score further decrease to 87.73% and 86.49%, yielding reductions of 2.72 and 3.94 percentage points relative to the full model. When the backbone is replaced with a Transformer, the two metrics further decrease to 83.64% and 83.23%. The relatively large performance gap suggests that the temporal feature extraction strategy is particularly important for this dataset. The results also indicate that anomaly injection provides additional robustness to the learned representations. For the Z24 dataset, the performance differences are smaller but remain consistent. The full model obtains 70.45% Accuracy and 70.43% F1-score. In contrast, removing anomaly injection decreases Accuracy and F1-score by 0.90 and 1.51 percentage points. Replacing the backbone with a Transformer results in Accuracy and F1-score of 69.09% and 68.91%, corresponding to decreases of 1.36 and 1.52 percentage points. Although the performance degradation is less pronounced than that on IASC-ASCE, neither modification improves the results.

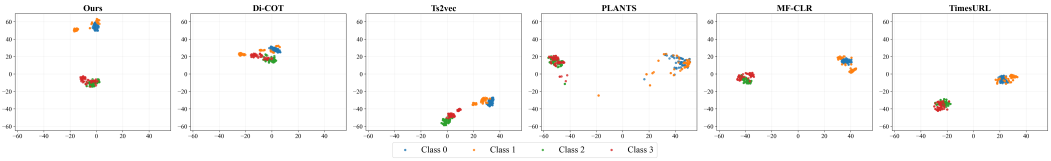
4.4 Representation visualization

To gain further insight into the discriminative quality of the learned embeddings, t-SNE visualizations of the representations produced by the proposed method and the competing baselines are shown in Figure 2 for the Steel Frame, IASC-ASCE, and Z24 datasets. Each point corresponds to a test

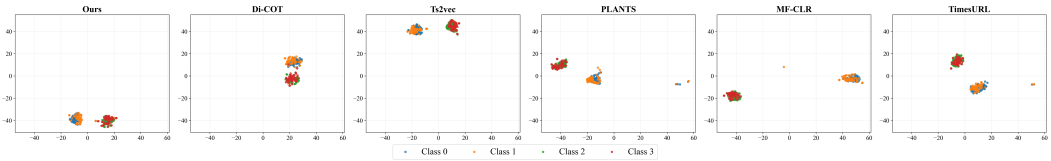
sample, colored by its structural health label. Well-separated and tightly clustered groups indicate that the learned representation captures class-specific vibration patterns effectively.



(a) Steel Frame



(b) IASC-ASCE



(c) Z24

Figure 2: t-SNE visualizations of the learned representations on the three datasets.

Across all three datasets, the proposed method consistently delivers the most favorable visualization: the embeddings are not only well-separated between classes but also compact within each class. This strong clustering structure corroborates the quantitative results and confirms that the framework learns representations that are both discriminative and generalizable. The baselines, while producing acceptable performance on some datasets, do not maintain the same level of consistency, which underlines the advantage of the proposed contrastive few-shot representation learning strategy for structural vibration data.

5 Conclusion

This study addresses the practical challenge of structural damage identification under degraded sensor data quality and proposes a data-quality-invariant time-series classification framework based on reconstruction-based contrastive representation learning. The proposed framework integrates anomaly injection, reconstruction-based contrastive learning, and few-shot prototype classification into a unified two-stage learning architecture. Extensive experiments were conducted on three structural vibration datasets, including the four-story steel frame, IASC-ASCE, and Z24 bridge datasets. Across all three datasets, the proposed framework consistently achieves the best performance in Accuracy, Precision, Recall, and F1-score, outperforming ten representative self-supervised time-series representation learning baselines. The consistent improvements across datasets and evaluation metrics demonstrate that the learned representations retain stronger class-discriminative information under degraded sensor conditions. Further ablation and sensitivity analyses indicate that both anomaly injection and the multi-scale temporal backbone contribute to the representation quality, while the latter has a comparatively greater impact on the final performance. In addition, t-SNE visualization reveals more compact intra-class distributions and clearer inter-class boundaries, providing complementary qualitative evidence for the superior discriminability of the learned representation space. Therefore, the proposed method is particularly suitable for practical structural health monitoring scenarios characterized by frequent sensor faults and degradation in data quality.

REFERENCES

Filippo Maria Bianchi, Lorenzo Livi, Karl Øyvind Mikalsen, Michael Kampffmeyer, and Robert Jenssen. Learning representations of multivariate time series with missing data. *Pattern Recognition*, 96:106973, 2019.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pp. 1597–1607. PMLR, 2020.

Zahra Zamanzadeh Darban, Geoffrey I Webb, Shirui Pan, Charu C Aggarwal, and Mahsa Salehi. CARLA: Self-supervised contrastive representation learning for time series anomaly detection. *Pattern Recognition*, 157:110874, 2025.

Jiaxiang Dong, Haixu Wu, Haoran Zhang, Li Zhang, Jianmin Wang, and Mingsheng Long. SimMTM: A simple pre-training framework for masked time-series modeling. *Advances in Neural Information Processing Systems*, 36:29996–30025, 2023.

Jufang Duan, Wei Zheng, Yangzhou Du, Wenfa Wu, Haipeng Jiang, and Hongsheng Qi. MF-CLR: Multi-frequency contrastive learning representation for time series. In *ICML*, pp. 11918–11939, 2024.

Jean-Yves Franceschi, Aymeric Dieuleveut, and Martin Jaggi. Unsupervised scalable representation learning for multivariate time series. *Advances in neural information processing systems*, 32, 2019.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8): 1735–1780, 1997.

Hassan Ismail Fawaz, Benjamin Lucas, Germain Forestier, Charlotte Pelletier, Daniel F Schmidt, Jonathan Weber, Geoffrey I Webb, Lhassane Idoumghar, Pierre-Alain Muller, and François Petitjean. InceptionTime: Finding AlexNet for time series classification. *Data mining and knowledge discovery*, 34(6):1936–1962, 2020.

Sean Kulinski, Saurabh Bagchi, and David I Inouye. Feature shift detection: Localizing which features have shifted via conditional distribution tests. *Advances in neural information processing systems*, 33:19523–19533, 2020.

Xiang Lan, Hanshu Yan, Shenda Hong, and Mengling Feng. Towards enhancing time series contrastive learning: A dynamic bad pair mining approach. In *International conference on learning representations*, volume 2024, pp. 10359–10379, 2024.

Seunghan Lee, Taeyoung Park, and Kibok Lee. Learning to embed time series patches independently. In *International Conference on Learning Representations*, volume 2024, pp. 27599–27620, 2024a.

Seunghan Lee, Taeyoung Park, and Kibok Lee. Soft contrastive learning for time series. In *International Conference on Learning Representations*, volume 2024, pp. 46815–46839, 2024b.

Jianbo Li, Lecheng Zheng, Yada Zhu, and Jingrui He. Outlier impact characterization for time series data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 11595–11603, 2021.

Jiexi Liu and Songcan Chen. TimesURL: Self-supervised contrastive learning for universal time series representation learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 13918–13926, 2024.

Yu Liu, Haoqin Yang, Jinping Sui, Hui Wang, Haipeng Li, Weimin Wang, and Qi Jia. Time series class-incremental learning via confidence-guided mask distillation and prototype-guided contrastive learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 23917–23925, 2026.

Lajanugen Logeswaran and Honglak Lee. An efficient framework for learning sentence representations. *arXiv preprint arXiv:1803.02893*, 2018.

Timo Milbich, Omair Ghori, Ferran Diego, and Bjoern Ommer. Unsupervised representation learning by discovering reliable image relations. *Pattern Recognition*, 102:107107, 2020.

Kevin Ni, Nithya Ramanathan, Mohamed Nabil Hajj Chehade, Laura Balzano, Sheela Nair, Sadaf Zahedi, Eddie Kohler, Greg Pottie, Mark Hansen, and Mani Srivastava. Sensor network data fault types. *ACM Transactions on Sensor Networks (TOSN)*, 5(3):1–29, 2009.

Yuqi Nie, Nam H Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. *arXiv preprint arXiv:2211.14730*, 2022.

Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

Cheng Pan, Woqin Luo, Fengzong Gong, and Ye Xia. Transfer learning-based open-set data anomaly detection for structural health monitoring practice. *Structural Health Monitoring*, pp. 14759217251411219, 2026.

Abdul-Kazeem Shamba, Kerstin Bach, and Gavin Taylor. Divide and contrast: Learning robust temporal features without augmentation. *arXiv preprint arXiv:2605.21241*, 2026.

Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. *Advances in neural information processing systems*, 30, 2017.

Sana Tonekaboni, Danny Eytan, and Anna Goldenberg. Unsupervised representation learning for time series with temporal neighborhood coding. *arXiv preprint arXiv:2106.00750*, 2021.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Jia Wang, Xiao Wang, and Chi Zhang. Plants: Periodicity-aware latent-state representation learning for multivariate time series. *arXiv preprint arXiv:2509.05478*, 2025.

Yucheng Wang, Yuecong Xu, Jianfei Yang, Min Wu, Xiaoli Li, Lihua Xie, and Zhenghua Chen. Graph-aware contrasting for multivariate time-series classification. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 15725–15734, 2024.

Gerald Woo, Chenghao Liu, Doyen Sahoo, Akshat Kumar, and Steven Hoi. CoST: Contrastive learning of disentangled seasonal-trend representations for time series forecasting. *arXiv preprint arXiv:2202.01575*, 2022.

Hongzuo Xu, Yijie Wang, Guansong Pang, Songlei Jian, Ning Liu, and Yongjun Wang. Rosas: Deep semi-supervised anomaly detection with contamination-resilient continuous supervision. *Information Processing & Management*, 60(5):103459, 2023.

Maxwell Xu, Alexander Moreno, Hui Wei, Benjamin M Marlin, and James Rehg. REBAR: Retrieval-based reconstruction for time-series contrastive learning. In *International Conference on Learning Representations*, volume 2024, pp. 15440–15464, 2024.

Qiantong Xu, Alexei Baevski, Tatiana Likhomanenko, Paden Tomasello, Alexis Conneau, Ronan Collobert, Gabriel Synnaeve, and Michael Auli. Self-training and pre-training are complementary for speech recognition. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3030–3034. IEEE, 2021.

Yiyuan Yang, Chaoli Zhang, Tian Zhou, Qingsong Wen, and Liang Sun. Dcdetector: Dual attention contrastive representation learning for time series anomaly detection. In *Proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining*, pp. 3033–3045, 2023.

Zhihan Yue, Yujing Wang, Juanyong Duan, Tianmeng Yang, Congrui Huang, Yunhai Tong, and Bixiong Xu. TS2Vec: Towards universal representation of time series. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pp. 8980–8987, 2022.

George Zerveas, Srideepika Jayaraman, Dhaval Patel, Anuradha Bhamidipaty, and Carsten Eickhoff. A transformer-based framework for multivariate time series representation learning. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, pp. 2114–2124, 2021.

Xu Zheng, Tianchun Wang, Wei Cheng, Aitian Ma, Haifeng Chen, Mo Sha, and Dongsheng Luo. Parametric augmentation for time series contrastive learning. In *International Conference on Learning Representations*, volume 2024, pp. 46906–46931, 2024.