Multi-View Contrastive Learning for Robust Domain Adaptation in Medical Time Series Analysis

YongKyung Oh YONGKYUNGOH@MEDNET.UCLA.EDU
Medical & Imaging Informatics (MII), UCLA

Alex A. T. Bui* BUIA@MII.UCLA.EDU
Medical & Imaging Informatics (MII), UCLA

Abstract

Adapting machine learning models to medical time series across different domains remains a challenge due to complex temporal dependencies and dynamic distribution shifts. Current approaches often focus on isolated feature representations, limiting their ability to fully capture the intricate temporal dynamics necessary for robust domain adaptation. In this work, we propose a novel framework leveraging multi-view contrastive learning to integrate temporal patterns, derivative-based dynamics, and frequency-domain features. Our method employs independent encoders and a hierarchical fusion mechanism to learn feature-invariant representations that are transferable across domains while preserving temporal coherence. Extensive experiments on diverse medical datasets, including electroencephalogram (EEG), electrocardiogram (ECG), and electromyography (EMG) demonstrate that our approach significantly outperforms state-of-the-art methods in transfer learning tasks. By advancing the robustness and generalizability of machine learning models, our framework offers a practical pathway for deploying reliable AI systems in diverse healthcare settings.

Data and Code Availability This study uses publicly available datasets in medical and healthcare domains, including SleepEEG (Kemp et al., 2000) and ECG (Clifford et al., 2017) for pre-training, and Epilepsy (Andrzejak et al., 2001), FD (Lessmeier et al., 2016), Gesture (Liu et al., 2009), and EMG (Goldberger et al., 2000) for fine-tuning. The datasets used in this study are publicly accessible via their respective repositories, with detailed documentation included in the supplementary material. Additionally, implementation details and code repository[1] are provided to facilitate reproducibility.

Institutional Review Board (IRB) This research uses publicly available datasets that do not include identifiable personal information and, as such, does not require IRB approval.

1. Introduction

Time series data form the backbone of modern healthcare monitoring and clinical decision-making, where the analysis of temporal patterns from diverse medical signals directly influences patient care outcomes and treatment strategies. However, effective analysis of time series data poses significant challenges due to complex temporal dependencies, non-stationarity, irregularity, and distribution shifts that occur across different domains (Hyndman, 2018; Hamilton, 2020; Oh et al., 2024c). These challenges are further amplified when models trained in one domain are applied to another, leading to degraded performance—a phenomenon known as *domain shift* (Li et al., 2017; Quiñonero-Candela et al., 2022).

Domain adaptation aims to address this issue by transferring knowledge learned from a large source domain to a small target domain, thereby enhancing model performance on the target data (Pan and Yang, 2009; Farahani et al., 2021). While domain adaptation has seen considerable success in fields like computer vision and natural language processing (Ganin and Lempitsky, 2015; Tzeng et al., 2017), its application to time series data remains less explored due to the inherent complexities of temporal information, including the alignment of temporal structures

* Corresponding Author

1. https://github.com/yongkyung-oh/Multi-View_Contrastive_Learning

and the mitigation of shifts in both feature and label distributions (Wilson et al., 2020; Shi et al., 2022).

Traditional domain adaptation techniques often fall short in addressing the intricate temporal dependencies and dynamic distribution shifts characteristic of time series data, frequently resulting in suboptimal performance. Approaches that focus exclusively on minimizing statistical distances between source and target domains, such as Maximum Mean Discrepancy (MMD)-based methods, tend to neglect the temporal relationships essential for effective modeling and knowledge transfer (Long et al., 2015; Cai et al., 2021; Liu and Xue, 2021). To address these challenges, recent advancements in contrastive learning and self-supervised frameworks have shown promise by capturing invariant features through representations based on temporal and frequency information (Eldele et al., 2021; Ozyurt et al., 2023; Zhang et al., 2024). Specifically, techniques such as feature alignment and adversarially-learned embeddings have demonstrated improved robustness against domain shifts, enhancing the applicability of these models to real-world scenarios (Jin et al., 2022; Ragab et al., 2023; Oh et al., 2023).

Despite these advancements, existing methods predominantly focus on a single type of feature representation or decomposition, limiting their ability to fully capture the multifaceted temporal characteristics of time series data (Zerveas et al., 2021; Cai et al., 2021; Yue et al., 2022; Lee et al., 2024). Furthermore, many approaches inadequately address the need for invariant representations that simultaneously preserve temporal coherence, transfer knowledge, and adapt to dynamic distribution shifts (Sun and Saenko, 2016; Weber et al., 2021; Oh et al., 2024b).

To further ground our research in practical and impactful applications, our study focuses on the medical domain, where time series data play a pivotal role. Medical time series, such as electrocardiograms (ECGs), electroencephalograms (EEGs), electronic health records (EHRs), and patient monitoring data often exhibit complex temporal dependencies and significant distribution shifts due to variations in patient populations, equipment, and clinical settings (Shickel et al., 2017; Purushotham et al., 2018; Harutyunyan et al., 2019; Xie et al., 2022). These shifts pose barriers to deploying machine learning models across diverse healthcare environments. Existing domain adaptation techniques are particularly limited in this context due to the need for preserving temporal coherence while addressing sensitive and dynamic distribution shifts.

Here, we propose a novel self-supervised framework for time series feature adaptation that leverages complementary representations learned through multi-view contrastive learning. Specifically, our approach integrates information from three distinct features: the *temporal-feature*, which captures inherent patterns and trends within the raw data; the *derivative-feature*, which characterizes local dynamics and trend changes over time; and the *frequency-feature*, which provides insights into global spectral features of the time series. By combining these complementary features, our framework aims to learn *domain-invariant representations* that are robust to distribution shifts while preserving temporal coherence. Extensive experiments on benchmark datasets demonstrate the superiority of our approach over state-of-the-art methods, validating the advantages of integrating complementary representations and leveraging contrastive learning.

The remainder of the paper is organized as follows. Section 2 reviews related work in time series domain adaptation and contrastive learning approaches. Section 3 introduces our proposed framework, detailing the multi-view feature extraction and hierarchical fusion mechanisms. Section 4 presents comprehensive experimental evaluations on medical time series datasets including 'one-to-one scenario' and 'one-to-many scenario'. Section 5 examines practical implications and limitations, while Section 6 outlines conclusions and future research directions.

2. Related works

Time series domain adaptation has been extensively studied as a means to address distribution shifts between source and target domains (Ben-David et al., 2010; Purushotham et al., 2017; Ragab et al., 2023; Zhang et al., 2024). Various recent approaches have been explored for time series domain adaptation, including sparse associative structure alignment for inter-variable relationships, contrastive learning for temporal and contextual patterns, and consistency enforcement between time and frequency domains (Cai et al., 2021; Eldele et al., 2021; Zhang et al., 2022). These methods highlight the importance of leveraging multiple feature representations to effectively address domain shifts in time series data.

Contrastive learning has emerged as a powerful tool for self-supervised representation learning, show-

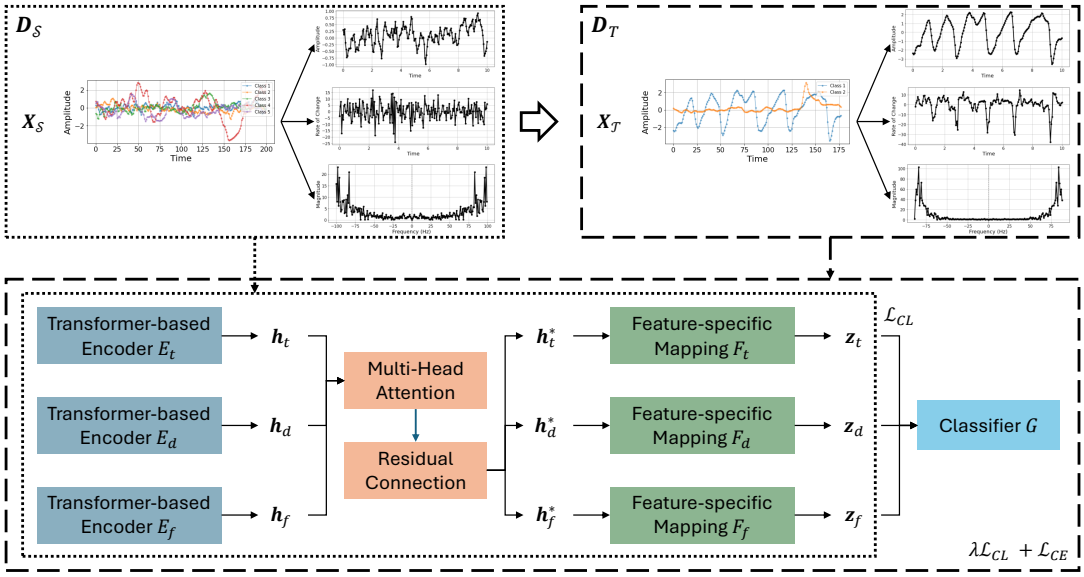


Figure 1: Overview of the proposed framework. The dotted box represents the multi-view contrastive pre-training phase using $\mathcal{D}_S$, while the dashed box corresponds to the fine-tuning phase using $\mathcal{D}_T$.

ing impressive results across various domains (Chen et al., 2020; He et al., 2020; Tian et al., 2020). In time series analysis, contrastive methods have been employed to learn robust representations by defining positive and negative pairs based on temporal proximity or data augmentations (Woo et al., 2022; Zhang et al., 2022). These approaches highlight the potential of leveraging invariances within temporal data to improve generalization across domains.

In the medical domain, domain adaptation for time series faces unique challenges. Medical time series datasets are characterized by high inter-patient variability, non-stationarity due to physiological conditions, and differences in data collection settings (He et al., 2023; Wang et al., 2024). These factors exacerbate distribution shifts and complicate the adaptation process. Approaches like domain adversarial networks and hierarchical feature alignment have been applied to mitigate these challenges, but often fail to fully address the critical need for preserving medical context and interpretability.

Still, existing methods often focus on a single type of feature or decomposition, limiting their ability to fully capture the inherent temporal characteristics of time series data. They may overlook the benefits of integrating complementary representations from multiple domains—such as time, derivative, and frequency domains—which can provide a more comprehensive understanding of the data. By focusing on these limitations, our study contributes to bridging the gap between robust representation learning and effective domain adaptation for time series in both general and medical contexts.

3. Methodology

3.1. Problem Formulation

Let $\mathcal{D}_\mathcal{S} = \{(\boldsymbol{X}_\mathcal{S}^{(i)}, y_\mathcal{S}^{(i)})\}_{i=1}^{N_\mathcal{S}}$ denote the source domain dataset, where $\boldsymbol{X}_\mathcal{S}^{(i)} \in \mathbb{R}^{T_\mathcal{S} \times d_\mathcal{S}}$ is a multivariate time series sample with $T_\mathcal{S}$ timesteps and $d_\mathcal{S}$ variables, and $y_\mathcal{S}^{(i)} \in \mathcal{Y}_\mathcal{S}$ is the corresponding label. Similarly, let $\mathcal{D}_\mathcal{T} = \{\boldsymbol{X}_\mathcal{T}^{(i)}, y_\mathcal{T}^{(i)}\}_{i=1}^{N_\mathcal{T}}$ denote the target domain dataset, where $\boldsymbol{X}_\mathcal{T}^{(i)} \in \mathbb{R}^{T_\mathcal{T} \times d_\mathcal{T}}$ and $y_\mathcal{T}^{(i)} \in \mathcal{Y}_\mathcal{T}$. Our goal is to learn encoders $E(\cdot)$, feature extractor $F(\cdot)$ and a classifier $G(\cdot)$ such that the classifier $G(F(E(\boldsymbol{X}_\mathcal{T})))$ performs well on the target domain, despite the distribution shift between $\mathcal{D}_\mathcal{S}$ and $\mathcal{D}_\mathcal{T}$.

Note that for contrastive pre-training, we treat the source domain $\mathcal{D}_\mathcal{S}$ as unlabeled, and thus the labels $y_\mathcal{S}^{(i)}$ are not utilized during representation learning.

3.2. Overview of the Proposed Framework

As shown in Figure 1, our framework consists of two phases: multi-view contrastive pre-training and fine-

tuning. In the pre-training phase, we leverage three complementary views of time series data from source domain $\mathcal{D}_S$: temporal, derivative, and frequency domains. Each view is processed through its respective encoder (E_t, E_d, E_f) and projector (F_t, F_d, F_f) to obtain view-specific representations in the contrastive learning space. These representations are optimized using contrastive loss $\mathcal{L}_{CL}$ to capture the inherent relationships across different views. During the fine-tuning phase, we transfer the pre-trained encoders and projectors to the target domain $\mathcal{D}_T$ and introduce a classification head G that maps the concatenated representations to the target task output, optimized using Cross-entropy loss $\mathcal{L}_{CE}$. The detailed components and methodology of our framework are elaborated in the following sections.

3.3. Multi-view Feature Extraction

3.3.1. Temporal-feature Encoder

The temporal-feature encoder $E_t(\cdot)$ captures inherent patterns and trends within the raw time series data. The temporal-feature representation is obtained as:

$$\boldsymbol{h}_t = E_t(\boldsymbol{X}) \in \mathbb{R}^{L\times D},$$

where L represents the length of time series and can be $T_{\mathcal{S}}$ or $T_{\mathcal{T}}$, and D is the hidden dimension.

3.3.2. Derivative-feature Encoder

To capture local dynamics and shifts in trends over time, we compute derivative features using finite difference interpolation. Given the time series $\boldsymbol{X}$, we approximate the first-order derivative $\dot{\boldsymbol{X}}$ as follows:

$$\dot{\boldsymbol{X}}_t = \frac{3\boldsymbol{X}_t - 4\boldsymbol{X}_{t-1} + \boldsymbol{X}_{t-2}}{2\Delta t},$$

where Δt is the time interval between consecutive samples, and $t = 2, \ldots, T$. An interpolation technique is used for derivative computation due to its smoothness properties and robustness to noise, as discussed in Morrill et al. (2022). The derivative-feature representation is then obtained by passing $\dot{\boldsymbol{X}}$ through the encoder $E_d(\cdot)$ and padding:

$$\boldsymbol{h}_d = E_d(\dot{\boldsymbol{X}}) \in \mathbb{R}^{L\times D}.$$

3.3.3. Frequency-feature Encoder

The frequency-feature encoder $E_f(\cdot)$ captures global spectral characteristics of the time series. We compute the frequency-feature representation using the Fast Fourier Transform (FFT). We take the magnitude of the FFT coefficients to obtain the amplitude spectrum:

$$\boldsymbol{X}_{\text{freq}} = |\text{FFT}(\boldsymbol{X})|.$$

The frequency-feature representation is then extracted using the encoder $E_f(\cdot)$:

$$\boldsymbol{h}_f = E_f(\boldsymbol{X}_{\text{freq}}) \in \mathbb{R}^{L\times D}.$$

We implement E_t, E_d, and E_f, using a Transformer encoder architecture (Vaswani, 2017), which effectively models long-range dependencies through self-attention mechanisms. By utilizing these encoders, our framework extracts rich features from each feature, which are later fused to form a comprehensive representation of the time series data.

3.4. Hierarchical Feature Fusion

To integrate the representations from the three features—time, derivative, and frequency—we employ a hierarchical fusion mechanism designed to capture complex interactions among these features. Although these features are derived from the same input data, the fusion process can enhance mutual information by integrating complementary representations (Oh et al., 2024a; Qiu et al., 2024). This fusion operates by applying multi-head attention across the stacked representations, allowing the model to focus on salient features and interactions between different features, discussed in Section 3.3.

Given the multi-view feature representations using encoders, $\boldsymbol{h}_t, \boldsymbol{h}_d, \boldsymbol{h}_f \in \mathbb{R}^{L\times D}$, where L is the sequence length, and D is the hidden dimension, we stack the three representations along a new dimension:

$$\boldsymbol{H} = \text{stack}(\boldsymbol{h}_t, \boldsymbol{h}_d, \boldsymbol{h}_f) \in \mathbb{R}^{L\times 3\times D}.$$

We then apply multi-head attention (MHA) on the stacked feature $\boldsymbol{H}$, so that the attention is computed across the feature dimension. After the attention mechanism, a residual connection and subsequent layer normalization are applied:

$$\boldsymbol{H}_{\text{output}} = \text{LayerNorm}(\boldsymbol{H} + \text{MHA}(\boldsymbol{H})).$$

We extract the updated representations for each feature by splitting $\boldsymbol{H}_{\text{output}}$ along the feature dimension:

$$\boldsymbol{H}_{\text{output}} = [\boldsymbol{h}_t^*, \boldsymbol{h}_d^*, \boldsymbol{h}_f^*],$$

where $\boldsymbol{h}_t^*, \boldsymbol{h}_d^*, \boldsymbol{h}_f^* \in \mathbb{R}^{L\times D}$ are the updated feature representations. These representations now contain information not only from their respective features but also from interactions with the other features.

To obtain the final representation for downstream task, we process each updated feature representation through a feature-specific mapping $F_k(\cdot)$, which are learnable feed-forward networks and averaging layer:

$$\boldsymbol{z}_k = F_k(\boldsymbol{h}_k^*) \quad \text{for} \quad k \in \{t, d, f\}.$$

By applying attention across stacked representations, the model learns to capture dependencies and interactions between different features. Additionally, the attention mechanism allows the model to focus on informative features between each other, improving the quality of the fused representation.

3.5. Optimization with Combined Losses

Our proposed framework leverages two key loss functions—feature-specific contrastive losses and supervised classification loss—optimized in two sequential stages: pre-training on the source dataset and fine-tuning on the target dataset.

Feature-Specific Contrastive Losses To enhance the quality of feature representations, we employ a contrastive loss for each feature type (temporal, derivative, and frequency). Positive and negative pairs are constructed using time series-specific augmentations as suggested by Zhang et al. (2022). For each feature $k \in \{t, d, f\}$, let $\boldsymbol{z}_k$ and $\tilde{\boldsymbol{z}}_k$ represent the original and augmented feature representations, respectively. The contrastive InfoNCE is defined as:

$$\mathcal{L}_{\text{CL}}^k = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{\exp(\text{sim}(\boldsymbol{z}_k^{(i)}, \tilde{\boldsymbol{z}}_k^{(i)})/\tau)}{\sum_{j=1}^{N} \exp(\text{sim}(\boldsymbol{z}_k^{(i)}, \tilde{\boldsymbol{z}}_k^{(j)})/\tau)},$$

where $\text{sim}(\boldsymbol{u}, \boldsymbol{v}) = \frac{\boldsymbol{u}^\top \boldsymbol{v}}{\|\boldsymbol{u}\|\|\boldsymbol{v}\|}$ is the cosine similarity, τ is the temperature parameter set to 0.07, as suggested by He et al. (2020), and N is the batch size. The total contrastive loss is the sum of feature-specific losses:

$$\mathcal{L}_{\text{CL}} = \sum_{k\in\{t,d,f\}} \mathcal{L}_{\text{CL}}^k.$$

Supervised Classification Loss For labeled target data, we use the supervised cross-entropy loss to train the classifier. Let $\boldsymbol{z}_{\text{combined}} = [\boldsymbol{z}_t, \boldsymbol{z}_d, \boldsymbol{z}_f]$ be the concatenated feature representations. For the classifier $G(\cdot)$, the predicted label for the i-th sample is:

$$\hat{y}^{(i)} = G(\boldsymbol{z}_{\text{combined}}^{(i)}).$$

The supervised classification loss is computed as:

$$\mathcal{L}_{\text{CE}} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} \mathbb{I}_{[y^{(i)}=c]} \log p_c^{(i)},$$

where C is the number of classes, $\mathbb{I}_{[\cdot]}$ is the indicator function, and $p_c^{(i)}$ denotes the predicted probability for class c.

Optimization Procedure The training process involves two stages:

1. **Pre-training on Source Dataset**: The model is trained on the source dataset using only the InfoNCE loss to learn feature-invariant representations across $\{t, d, f\}$. This stage optimizes the encoders E_k and projection heads F_k for each feature type $k \in \{t, d, f\}$, ensuring that each feature view preserves its intra-view information independently through contrastive learning.

2. **Fine-tuning on Target Dataset**: The pre-trained model is fine-tuned on the target dataset using a weighted combination of the InfoNCE loss and the cross-entropy loss. This stage trains the encoders E_k, projection heads F_k, and the classifier G:

$$\mathcal{L}_{\text{total}} = \frac{1}{N}\sum_{i=1}^{N}\left(\lambda \mathcal{L}_{\text{CL}}^{(i)} + \mathcal{L}_{\text{CE}}^{(i)}\right),$$

where λ is a hyperparameter balancing the two losses, set to 0.1. This fine-tuning phase leverages hierarchical fusion to integrate multi-view information, capturing inter-view dependencies that enhance predictive performance.

The losses are optimized using the Adam optimizer (Kingma and Ba, 2014), ensuring efficient convergence during both stages. The rationale behind our architectural design and its theoretical foundation are provided in the Appendix A.

4. Experiment

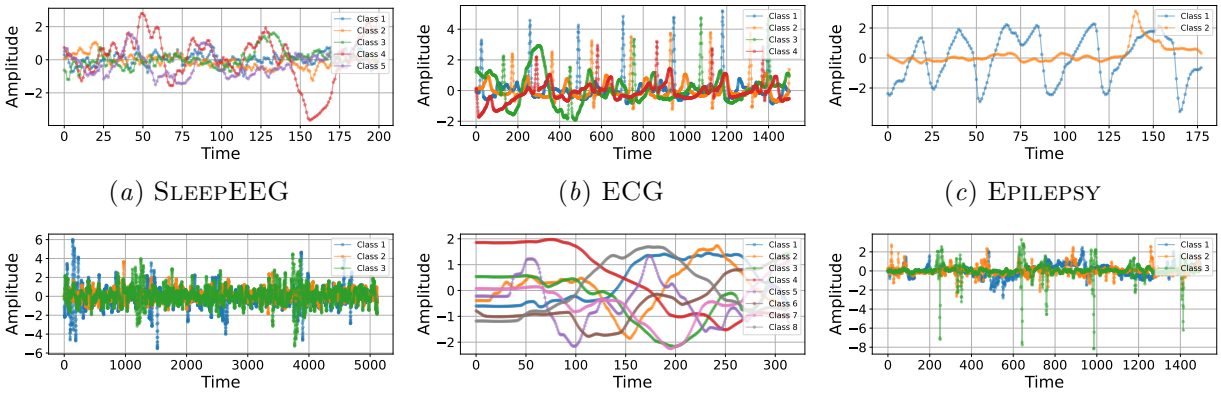
4.1. Experimental Setting

We evaluate our framework following the experimental pipeline introduced in Zhang et al. (2022) and Dong et al. (2024). Preprocessed data can be accessed through corresponding repository[2]. We utilize

2. https://github.com/mims-harvard/TFC-pretraining

Table 1: Description of dataset statistics. For the number of samples in the target (fine-tuning) dataset, "n_1 / n_2 / n_3" indicates n_1 samples for fine-tuning, n_2 samples for validation, and n_3 samples for testing.

Dataset	# Samples	# Channels	# Classes	Length	Freq (Hz)
SleepEEG	371,055	1	5	200	100
ECG	43,673	1	4	1,500	300
Epilepsy	60 / 20 / 11,420	1	2	178	174
FD	60 / 21 / 13,559	1	3	5,120	64K
Gesture	320 / 120 / 120	3	8	315	100
EMG	122 / 41 / 41	1	3	1,500	4,000



(*a*) SleepEEG (*b*) ECG (*c*) Epilepsy

(*d*) FD (*e*) Gesture (*f*) EMG

Figure 2: Representative samples from each dataset, with different colors indicating distinct classes

SleepEEG as the source domain for pre-training, chosen for its rich temporal patterns and substantial data volume. The pre-trained model is then fine-tuned separately on four target domains: Epilepsy, FD, Gesture, and EMG. Additional to that, we use ECG as source dataset for sensitivity analysis.

Our experimental setup allows us to assess the model's transferability across diverse domains characterized by varying sampling rates, channel dimensions, sequence durations, and class compositions.

Data Statistics of each dataset are described in Table 1. Representative samples from each dataset are illustrated in Figure 2, with comprehensive dataset descriptions provided in the Appendix B.

4.2. Benchmark Methods

Various approaches have been developed for time series representation learning, each with distinct characteristics. **TST** (Zerveas et al., 2021) uses a Transformer-based framework for multivariate time series. **TS-SD** (Shi et al., 2021) applies self-supervised learning with sliding windows and dynamic time warping. **SimCLR** (Tang et al., 2020) adapts a prominent contrastive learning framework (Chen et al., 2020) from computer vision to time series. **TS-TCC** (Eldele et al., 2021) leverages temporal and contextual contrasting. **CLOCS** (Kiyasseh et al., 2021) focuses on contrastive learning for cardiac signals in the medical domain. **Mixing-up** (Wickstrøm et al., 2022) utilizes time series mixing for data augmentation. **TS2Vec** (Yue et al., 2022) provides a universal framework using hierarchical contrastive loss. **TF-C** (Zhang et al., 2022) emphasizes temporal-frequency contrasting for self-supervised learning. **CoST** (Woo et al., 2022) disentangles seasonal-trend representations through contrastive learning. **Ti-MAE** (Li et al., 2023) employs a masked autoencoder for self-supervised learning. **SimMTM** (Dong et al., 2024) uses masked time points with weighted neighbor aggregation for representation learning.

For the fair performance comparisons, we used identical pipeline and reported performance metrics from Zhang et al. (2022) and Dong et al. (2024).

Table 2: Performance comparison with benchmark method: SleepEEG → Epilepsy
(mean ± standard deviation over five runs; 0.000 indicates unreported variance in the referenced studies)

Method \ Metric	Accuracy	Precision	Recall	F1 score
TST (Zerveas et al., 2021)	$0.802_{\pm 0.000}$	$0.401_{\pm 0.000}$	$0.500_{\pm 0.000}$	$0.445_{\pm 0.000}$
TS-SD (Shi et al., 2021)	$0.895_{\pm 0.052}$	$0.802_{\pm 0.224}$	$0.765_{\pm 0.149}$	$0.777_{\pm 0.186}$
SimCLR (Tang et al., 2020)	$0.907_{\pm 0.034}$	$0.922_{\pm 0.017}$	$0.786_{\pm 0.107}$	$0.818_{\pm 0.100}$
TS-TCC (Eldele et al., 2021)	$0.925_{\pm 0.010}$	$0.945_{\pm 0.005}$	$0.818_{\pm 0.026}$	$0.863_{\pm 0.022}$
CLOCS (Kiyasseh et al., 2021)	$0.951_{\pm 0.003}$	$0.930_{\pm 0.007}$	$0.913_{\pm 0.017}$	$0.921_{\pm 0.007}$
Mixing-up (Wickstrøm et al., 2022)	$0.802_{\pm 0.000}$	$0.401_{\pm 0.000}$	$0.500_{\pm 0.000}$	$0.445_{\pm 0.000}$
TS2Vec (Yue et al., 2022)	$0.940_{\pm 0.004}$	$0.906_{\pm 0.012}$	$0.904_{\pm 0.012}$	$0.905_{\pm 0.007}$
TF-C (Zhang et al., 2022)	$0.950_{\pm 0.025}$	$\mathbf{0.946}_{\pm \mathbf{0.011}}$	$0.891_{\pm 0.022}$	$0.915_{\pm 0.053}$
CoST (Woo et al., 2022)	$0.884_{\pm 0.000}$	$0.882_{\pm 0.000}$	$0.723_{\pm 0.000}$	$0.769_{\pm 0.000}$
Ti-MAE (Li et al., 2023)	$0.897_{\pm 0.000}$	$0.724_{\pm 0.000}$	$0.675_{\pm 0.000}$	$0.686_{\pm 0.000}$
SimMTM (Dong et al., 2024)	$0.955_{\pm 0.000}$	$0.934_{\pm 0.000}$	$0.923_{\pm 0.000}$	$0.928_{\pm 0.000}$
Proposed method	$\mathbf{0.956}_{\pm \mathbf{0.002}}$	$0.936_{\pm 0.004}$	$\mathbf{0.935}_{\pm \mathbf{0.004}}$	$\mathbf{0.931}_{\pm \mathbf{0.003}}$

Table 3: Sensitivity analysis on Epilepsy with a variety of learning strategies: different source datasets (SleepEEG and ECG), self-supervised learning, and random initialization (without pre-training)
(For each experimental scenario, classifier $G(\cdot)$ was initialized and subsequently trained by target dataset. Default setting fine-tunes of all $E_k(\cdot)$, $F_k(\cdot)$, and $G(\cdot)$. 'freeze scenario' indicates that $E_k(\cdot)$ and $F_k(\cdot)$ are fixed after pre-training, while only $G(\cdot)$ are optimized during fine-tuning, where $k \in \{t, d, f\}$.)

Learning Strategy \ Metric	Accuracy	Precision	Recall	F1 score
Source: SleepEEG	$0.956_{\pm 0.002}$	$0.936_{\pm 0.004}$	$0.935_{\pm 0.004}$	$0.931_{\pm 0.003}$
– freeze $E_k(\cdot)$ and $F_k(\cdot)$	$0.952_{\pm 0.005}$	$0.925_{\pm 0.012}$	$0.931_{\pm 0.004}$	$0.924_{\pm 0.006}$
Source: ECG	$0.953_{\pm 0.002}$	$0.934_{\pm 0.004}$	$0.940_{\pm 0.008}$	$0.927_{\pm 0.003}$
– freeze $E_k(\cdot)$ and $F_k(\cdot)$	$0.952_{\pm 0.002}$	$0.932_{\pm 0.012}$	$0.935_{\pm 0.006}$	$0.924_{\pm 0.002}$
Source: Epilepsy	$0.953_{\pm 0.002}$	$0.931_{\pm 0.009}$	$0.941_{\pm 0.003}$	$0.926_{\pm 0.002}$
– freeze $E_k(\cdot)$ and $F_k(\cdot)$	$0.953_{\pm 0.003}$	$0.928_{\pm 0.012}$	$0.939_{\pm 0.004}$	$0.925_{\pm 0.002}$
Random initialization	$0.948_{\pm 0.004}$	$0.911_{\pm 0.014}$	$0.936_{\pm 0.006}$	$0.920_{\pm 0.004}$

Table 4: Comprehensive analysis of the proposed framework's components on Epilepsy
(1) hierarchical feature fusion, (2) fine-tuning loss configurations with λ, and (3) combinations of features $\{\boldsymbol{h}_t, \boldsymbol{h}_d, \boldsymbol{h}_f\}$. (The proposed method utilizes all three features with $\lambda = 0.1$ as the default configuration.)

Model Component \ Metric	Accuracy	Precision	Recall	F1 score
Proposed method	$0.956_{\pm 0.002}$	$0.936_{\pm 0.004}$	$0.935_{\pm 0.004}$	$0.931_{\pm 0.003}$
– use $\mathcal{L}_{CE}$ only	$0.955_{\pm 0.003}$	$0.930_{\pm 0.005}$	$0.935_{\pm 0.003}$	$0.929_{\pm 0.004}$
w/o feature fusion	$0.956_{\pm 0.002}$	$0.934_{\pm 0.007}$	$0.936_{\pm 0.003}$	$0.930_{\pm 0.004}$
– use $\mathcal{L}_{CE}$ only	$0.952_{\pm 0.006}$	$0.922_{\pm 0.015}$	$0.936_{\pm 0.003}$	$0.925_{\pm 0.007}$
$\{\boldsymbol{h}_t, \boldsymbol{h}_d\}$	$0.945_{\pm 0.003}$	$0.925_{\pm 0.011}$	$0.916_{\pm 0.006}$	$0.911_{\pm 0.004}$
$\{\boldsymbol{h}_t, \boldsymbol{h}_f\}$	$0.952_{\pm 0.003}$	$0.926_{\pm 0.012}$	$0.928_{\pm 0.010}$	$0.925_{\pm 0.003}$
$\{\boldsymbol{h}_d, \boldsymbol{h}_f\}$	$0.955_{\pm 0.003}$	$0.934_{\pm 0.006}$	$0.935_{\pm 0.005}$	$0.928_{\pm 0.004}$
$\boldsymbol{h}_t$ only	$0.931_{\pm 0.008}$	$0.890_{\pm 0.021}$	$0.898_{\pm 0.005}$	$0.892_{\pm 0.009}$
$\boldsymbol{h}_d$ only	$0.940_{\pm 0.002}$	$0.907_{\pm 0.004}$	$0.910_{\pm 0.004}$	$0.905_{\pm 0.002}$
$\boldsymbol{h}_f$ only	$0.955_{\pm 0.003}$	$0.929_{\pm 0.009}$	$0.936_{\pm 0.005}$	$0.930_{\pm 0.005}$

4.3. Results with One-to-one Scenario

In this scenario, the framework is pre-trained on the SleepEEG dataset and fine-tuned on the Epilepsy dataset, both comprising single-channel EEG signals. The SleepEEG dataset provides recordings focused on sleep stages, while the Epilepsy dataset involves seizure and non-seizure classification. Despite both datasets being single-channel EEG, they differ in key aspects, including the scalp electrode positions used for recording, the physiological phenomena tracked (sleep patterns versus epilepsy activity), and the patient populations involved in the data collection.

Table 2 demonstrates a detailed performance comparison of the proposed method against several benchmark approaches for domain adaptation, specifically transferring from SleepEEG to Epilepsy. The proposed method outperforms all benchmarks across accuracy, precision, recall, and F1 score, achieving an F1 score of 0.931 ± 0.003, indicating its robustness and superior ability for domain adaptation task in medical time series analysis.

Table 3 explores the sensitivity of the proposed method on the Epilepsy dataset with various learning strategies. Pre-training on the SleepEEG dataset yields the highest F1 score (0.931 ± 0.003) when using fine-tuning with trainable encoders. Freezing the encoders results in a slight decrease in performance (F1: 0.924 ± 0.004), demonstrating the importance of continued encoder updates during fine-tuning. Pre-training on the ECG dataset also achieves competitive results, indicating the general adaptability of the proposed method across pre-training sources. However, pre-training directly on Epilepsy, which is self-supervised manner, shows slightly lower performance, suggesting that leveraging a larger, more complex source dataset like SleepEEG provides better feature generalization for the target domain. Interestingly, even with random initialization, the proposed method achieves an F1 score of 0.920 ± 0.004, indicating the robustness of its hierarchical feature fusion mechanism. This robustness highlights the method's ability to perform well even in the absence of domain-specific pre-training, though optimal results are achieved with targeted pre-training strategies.

Table 4 examines the contributions of different components in the proposed framework on the Epilepsy dataset, focusing on hierarchical feature fusion, fine-tuning loss configurations, and feature view combinations. The full proposed method, utilizing all features ($\boldsymbol{h}_t$, $\boldsymbol{h}_d$, $\boldsymbol{h}_f$) with hierarchical fusion and fine-tuning loss ($\mathcal{L}_{\text{CE}}$), achieves the highest F1 score. The results across these settings collectively demonstrate the robustness, adaptability, and effectiveness of the proposed multi-view contrastive learning framework in addressing domain adaptation challenges in time series analysis.

4.4. Results with One-to-many Scenario

In the one-to-many evaluation, the framework is pre-trained on the SleepEEG dataset and subsequently fine-tuned independently on multiple target datasets: Epilepsy, FD, Gesture, and EMG. This approach leverages a single, well-pretrained model from SleepEEG, without restarting the pre-training process for each fine-tuning task. This evaluation tests the adaptability and generalizability of the proposed framework across datasets with diverse data types, collection protocols, and physiological or operational contexts. The fine-tuning stage involves task-specific adjustments while leveraging the feature-invariant representations learned during pre-training.

Table 5 presents the performance comparison across different methods, where the F1 score is used as the primary evaluation metric. Additional metrics, including accuracy, precision, and recall, are provided in the Appendix C for a more comprehensive assessment. Our method consistently outperforms the benchmark approaches, demonstrating its effectiveness in diverse experimental settings.

In Table 6, we present an evaluation comparing different data sources, self-supervised learning approaches, and random initialization strategies. For self-supervised learning, the target dataset was utilized during the pre-training phase. In contrast, for the random initialization approach, the pre-training phase was omitted, and fine-tuning was performed directly. Across various scenarios, our proposed method demonstrates consistent and robust performance overall. Notably, the performance on the Gesture dataset decreases in the absence of pre-training on a larger source dataset, highlighting the importance of a substantial pre-training phase. Conversely, the EMG dataset achieves improved results when pre-trained with the ECG, suggesting that the specific characteristics of the target data play a pivotal role in performance for this case.

Figure 3 provides a brief comparison of performance across different feature combinations using pre-trained model using SleepEEG. These findings

Table 5: F1 score comparison between our method and benchmark approaches. All models are pre-trained on SleepEEG dataset and evaluated through fine-tuning on four different target datasets. (mean ± standard deviation over five runs; 0.000 indicates unreported variance in the referenced studies)

Method \ Target Dataset	Epilepsy	FD	Gesture	EMG
TST (Zerveas et al., 2021)	$0.445_{\pm 0.000}$	$0.413_{\pm 0.000}$	$0.660_{\pm 0.000}$	$0.689_{\pm 0.000}$
TS-SD (Shi et al., 2021)	$0.777_{\pm 0.186}$	$0.570_{\pm 0.033}$	$0.666_{\pm 0.044}$	$0.211_{\pm 0.000}$
SimCLR (Tang et al., 2020)	$0.818_{\pm 0.100}$	$0.422_{\pm 0.114}$	$0.496_{\pm 0.187}$	$0.471_{\pm 0.149}$
TS-TCC (Eldele et al., 2021)	$0.863_{\pm 0.022}$	$0.542_{\pm 0.034}$	$0.698_{\pm 0.036}$	$0.590_{\pm 0.095}$
CLOCS (Kiyasseh et al., 2021)	$0.921_{\pm 0.007}$	$0.475_{\pm 0.049}$	$0.401_{\pm 0.060}$	$0.514_{\pm 0.041}$
Mixing-up (Wickstrøm et al., 2022)	$0.445_{\pm 0.000}$	$0.727_{\pm 0.023}$	$0.650_{\pm 0.031}$	$0.154_{\pm 0.020}$
TS2Vec (Yue et al., 2022)	$0.905_{\pm 0.007}$	$0.439_{\pm 0.011}$	$0.657_{\pm 0.039}$	$0.677_{\pm 0.050}$
TF-C (Zhang et al., 2022)	$0.915_{\pm 0.053}$	$0.749_{\pm 0.027}$	$0.757_{\pm 0.031}$	$0.768_{\pm 0.031}$
CoST (Woo et al., 2022)	$0.769_{\pm 0.000}$	$0.348_{\pm 0.000}$	$0.664_{\pm 0.000}$	$0.353_{\pm 0.000}$
Ti-MAE (Li et al., 2023)	$0.686_{\pm 0.000}$	$0.666_{\pm 0.000}$	$0.684_{\pm 0.000}$	$0.709_{\pm 0.000}$
SimMTM (Dong et al., 2024)	$0.928_{\pm 0.000}$	$0.751_{\pm 0.000}$	$0.787_{\pm 0.000}$	$\mathbf{0.981}_{\pm \mathbf{0.000}}$
Proposed method	$\mathbf{0.931}_{\pm \mathbf{0.003}}$	$\mathbf{0.867}_{\pm \mathbf{0.012}}$	$\mathbf{0.820}_{\pm \mathbf{0.006}}$	$0.977_{\pm 0.017}$

Table 6: F1 score comparison between our method with a variety of learning strategies: different source datasets (SleepEEG and ECG), self-supervised learning, and random initialization (without pre-training)

Learning Strategy \ Target Dataset	Epilepsy	FD	Gesture	EMG
Source: SleepEEG	$0.931_{\pm 0.003}$	$0.867_{\pm 0.012}$	$0.820_{\pm 0.006}$	$0.977_{\pm 0.017}$
Source: ECG	$0.927_{\pm 0.003}$	$0.878_{\pm 0.006}$	$0.808_{\pm 0.012}$	$0.992_{\pm 0.019}$
Self-supervised learning	$0.926_{\pm 0.002}$	$0.865_{\pm 0.008}$	$0.786_{\pm 0.006}$	$0.964_{\pm 0.011}$
Random initialization	$0.920_{\pm 0.004}$	$0.877_{\pm 0.006}$	$0.773_{\pm 0.008}$	$0.972_{\pm 0.019}$

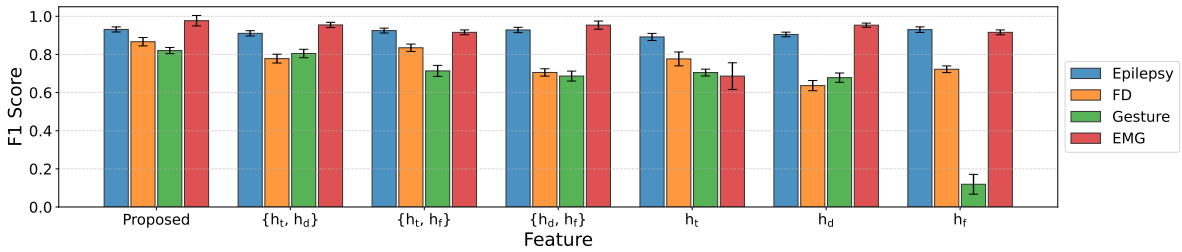


Figure 3: F1 score comparison of the proposed method with feature combinations using $\{\boldsymbol{h}_t, \boldsymbol{h}_d, \boldsymbol{h}_f\}$

underscore the critical role of selecting features that align closely with the inherent characteristics of the target dataset, as inappropriate feature emphasis can severely impair model effectiveness. The results reveal a notable decline in performance on the EMG dataset when temporal features are emphasized, while the Gesture dataset suffers a significant drop when frequency features are prioritized. Additional experimental results and analyses across target datasets can be found in the Appendix C.

5. Discussion

Effectiveness of proposed framework Our results demonstrate the substantial benefits of integrating complementary views (temporal, derivative, frequency) for robust domain adaptation. While our sensitivity analysis confirms that the proposed feature extraction and fusion mechanisms effectively learn transferable features, it also indicates that domain knowledge is crucial in choosing an optimal source dataset, as the best fit can vary by target task.

Feature Contribution Analysis The ablation studies provide valuable insights into the relative importance of different features. As demonstrated in Figure 3, while the combination of all three feature types ($\boldsymbol{h}_t$, $\boldsymbol{h}_d$, $\boldsymbol{h}_f$) generally yields the best performance, the relative importance varies across datasets.

Model Complexity We maintain three parallel Transformer-based encoders, one for each view. A standard m-layer Transformer has $\mathcal{O}(m\,L^2\,D)$ time complexity; as we instantiate three encoders, this adds a constant factor $\mathcal{O}(3 \times m\,L^2\,D)$ plus one multi-head attention step for fusion ($\mathcal{O}(L^2\,D)$). Nevertheless, it does not alter the overall $\mathcal{O}(m\,L^2\,D)$ scaling, meaning our method remains in the same computational class as a single Transformer. Moreover, in practical scenarios L is typically much larger than m, so the sequence-length term L^2 often dominates the runtime across most layers.

If L is large, efficient-attention mechanisms (Wang et al., 2020; Choromanski et al., 2021; Dao et al., 2022) can reduce $\mathcal{O}(L^2)$ to $\mathcal{O}(L \log L)$ or $\mathcal{O}(L)$, thus preserving tractability for real-world time series tasks. A promising future direction is to explore more efficient attention mechanisms and dynamic view selection to further reduce complexity.

Comprehensive Practical Analysis Despite the benefits of multi-view Transformer encoders, our design can increase computational requirement in both training and inference. To mitigate this, we employ automatic mixed precision (AMP) (Micikevicius et al., 2018), which reduces GPU memory requirements and accelerates large-batch contrastive pre-training. We also enable AMP at inference time, thereby improving end-to-end latency. Preliminary results are included in Appendix D.1.

In addition to that, real-world data often feature irregularly-sampled (Appendix D.2) or incomplete (Appendix D.3). We also explore the method's applicability in multivariate time series classification beyond domain adaptation (Appendix D.4). To evaluate robustness under these scenarios, we conduct additional analyses described in Appendix D, which provide in-depth discussions and empirical findings.

Note to Practitioner This work presents a multi-view contrastive learning framework for medical time series domain adaptation, addressing distribution shifts and temporal dependencies. Practitioners can leverage this methodology to enhance diagnostic and monitoring systems through robust transfer learning across diverse healthcare settings. By integrating temporal, derivative, and frequency features, the framework offers a practical solution for limited labeled data and variable collection protocols, providing actionable insights for deploying reliable machine learning models in clinical and public health contexts.

6. Conclusion

We proposed a novel framework for time series domain adaptation, leveraging multi-view contrastive learning to integrate complementary representations from temporal, derivative, and frequency features. By employing independent encoders and a hierarchical fusion mechanism, the framework effectively captures complex temporal dynamics and learns robust, feature-invariant representations that are transferable across domains. Extensive experiments on benchmark datasets demonstrate that the proposed method consistently outperforms state-of-the-art approaches, highlighting the benefits of multi-view feature integration for addressing domain adaptation challenges in time series analysis.

Additionally, incorporating domain-specific prior knowledge into the feature extraction process may further enhance model performance. Future work should investigate adaptive weighting mechanisms for feature fusion and explore the applicability of the framework to other types of medical time series data, including irregular or high-dimensional signals.

In the medical domain, this framework shows significant promise for improving diagnostic tools and patient monitoring systems. By learning robust representations, it addresses key challenges such as distribution shifts, data sparsity, and noise in medical time series data. The method's ability to transfer knowledge from large pre-training datasets to smaller, specific datasets facilitates applications in scenarios with limited labeled data—a common constraint in medical research.

Acknowledgments

This research was partially supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (RS-2024-00407852), and Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health and Welfare, Republic of Korea (HI19C1095).

References

Ralph G Andrzejak, Klaus Lehnertz, Florian Mormann, Christoph Rieke, Peter David, and Christian E Elger. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E*, 64(6):061907, 2001.

Anthony Bagnall, Hoang Anh Dau, Jason Lines, Michael Flynn, James Large, Aaron Bostrom, Paul Southam, and Eamonn Keogh. The uea multivariate time series classification archive, 2018. *arXiv preprint arXiv:1811.00075*, 2018.

Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. *Machine learning*, 79:151–175, 2010.

Ruichu Cai, Jiawei Chen, Zijian Li, Wei Chen, Keli Zhang, Junjian Ye, Zhuozhang Li, Xiaoyan Yang, and Zhenjie Zhang. Time series domain adaptation via sparse associative structure alignment. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 6859–6867, 2021.

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

Krzysztof Marcin Choromanski, Valerii Likhosherstov, David Dohan, Xingyou Song, Andreea Gane, Tamas Sarlos, Peter Hawkins, Jared Quincy Davis, Afroz Mohiuddin, Lukasz Kaiser, David Benjamin Belanger, Lucy J Colwell, and Adrian Weller. Rethinking attention with performers. In *International Conference on Learning Representations*, 2021.

Gari D Clifford, Chengyu Liu, Benjamin Moody, H Lehman Li-wei, Ikaro Silva, Qiao Li, AE Johnson, and Roger G Mark. Af classification from a short single lead ecg recording: The physionet/computing in cardiology challenge 2017. In *2017 Computing in Cardiology (CinC)*, pages 1–4. IEEE, 2017.

Thomas M Cover. *Elements of information theory*. John Wiley & Sons, 1999.

Tri Dao, Dan Fu, Stefano Ermon, Atri Rudra, and Christopher Ré. Flashattention: Fast and memory-efficient exact attention with io-awareness. *Advances in neural information processing systems*, 35:16344–16359, 2022.

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, 2024.

Emadeldeen Eldele, Mohamed Ragab, Zhenghua Chen, Min Wu, Chee Keong Kwoh, Xiaoli Li, and Cuntai Guan. Time-series representation learning via temporal and contextual contrasting. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pages 2352–2359, 2021.

Abolfazl Farahani, Sahar Voghoei, Khaled Rasheed, and Hamid R Arabnia. A brief review of domain adaptation. *Advances in data science and information engineering: proceedings from ICDATA 2020 and IKE 2020*, pages 877–894, 2021.

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

Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In *International conference on machine learning*, pages 1180–1189. PMLR, 2015.

Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley. Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *circulation*, 101(23):e215–e220, 2000.

James D Hamilton. *Time series analysis*. Princeton university press, 2020.

Wei Han, Hui Chen, and Soujanya Poria. Improving multimodal fusion with hierarchical mutual information maximization for multimodal sentiment analysis. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9180–9192, 2021.

Shilei Hao, Zhihai Wang, Afanasiev D Alexander, Jidong Yuan, and Wei Zhang. Micos: Mixed supervised contrastive learning for multivariate time series classification. *Knowledge-Based Systems*, 260:110158, 2023.

Hrayr Harutyunyan, Hrant Khachatrian, David C Kale, Greg Ver Steeg, and Aram Galstyan. Multitask learning and benchmarking with clinical time series data. *Scientific data*, 6(1):96, 2019.

Huan He, Owen Queen, Teddy Koker, Consuelo Cuevas, Theodoros Tsiligkaridis, and Marinka Zitnik. Domain adaptation for time series under feature and label shifts. In *International Conference on Machine Learning*, pages 12746–12774. PMLR, 2023.

Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9729–9738, 2020.

Marcel Hirt, Domenico Campolo, Victoria Leong, and Juan-Pablo Ortega. Learning multi-modal generative models with permutation-invariant encoders and tighter variational objectives. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856.

RJ Hyndman. *Forecasting: principles and practice*. OTexts, 2018.

Xiaoyong Jin, Youngsuk Park, Danielle Maddix, Hao Wang, and Yuyang Wang. Domain adaptation for time series forecasting via attention sharing. In *International Conference on Machine Learning*, pages 10280–10297. PMLR, 2022.

Bob Kemp, Aeilko H Zwinderman, Bert Tuk, Hilbert AC Kamphuisen, and Josefien JL Oberye. Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the eeg. *IEEE Transactions on Biomedical Engineering*, 47(9):1185–1194, 2000.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Dani Kiyasseh, Tingting Zhu, and David A Clifton. Clocs: Contrastive learning of cardiac signals across space, time, and patients. In *International Conference on Machine Learning*, pages 5606–5615. PMLR, 2021.

Seunghan Lee, Taeyoung Park, and Kibok Lee. Soft contrastive learning for time series. In *The Twelfth International Conference on Learning Representations*, 2024.

Christian Lessmeier, James Kuria Kimotho, Detmar Zimmer, and Walter Sextro. Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification. In *PHM Society European Conference*, volume 3, 2016.

Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy M Hospedales. Deeper, broader and artier domain generalization. In *Proceedings of the IEEE international conference on computer vision*, pages 5542–5550, 2017.

Zhe Li, Zhongwen Rao, Lujia Pan, Pengyun Wang, and Zenglin Xu. Ti-mae: Self-supervised masked time series autoencoders. *arXiv preprint arXiv:2301.08871*, 2023.

Jiayang Liu, Lin Zhong, Jehan Wickramasuriya, and Venu Vasudevan. uwave: Accelerometer-based personalized gesture recognition and its applications. *Pervasive and Mobile Computing*, 5(6):657–675, 2009.

Qiao Liu and Hui Xue. Adversarial spectral kernel matching for unsupervised time series domain adaptation. In *IJCAI*, pages 2744–2750, 2021.

Mingsheng Long, Yue Cao, Jianmin Wang, and Michael Jordan. Learning transferable features with deep adaptation networks. In *International conference on machine learning*, pages 97–105. PMLR, 2015.

Paulius Micikevicius, Sharan Narang, Jonah Alben, Gregory Diamos, Erich Elsen, David Garcia, Boris Ginsburg, Michael Houston, Oleksii Kuchaiev, Ganesh Venkatesh, and Hao Wu. Mixed precision training. In *International Conference on Learning Representations*, 2018.

James Morrill, Patrick Kidger, Lingyi Yang, and Terry Lyons. On the choice of interpolation scheme for neural cdes. *Transactions on Machine Learning Research*, 2022(9), 2022.

YongKyung Oh, Juhui Lee, and Sungil Kim. Sensor drift compensation for gas mixture classification in batch experiments. *Quality and Reliability Engineering International*, 39(6):2422–2437, 2023.

YongKyung Oh, Sungil Kim, and Alex AT Bui. Deep interaction feature fusion for robust human activity recognition. In *International Joint Conference on Artificial Intelligence*, pages 99–116. Springer, 2024a.

YongKyung Oh, Chiehyeon Lim, Junghye Lee, Sewon Kim, and Sungil Kim. Multichannel convolution neural network for gas mixture classification. *Annals of Operations Research*, 339(1):261–295, 2024b.

YongKyung Oh, Dongyoung Lim, and Sungil Kim. Stable neural stochastic differential equations in analyzing irregular time series data. In *The Twelfth International Conference on Learning Representations*, 2024c.

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

Yilmazcan Ozyurt, Stefan Feuerriegel, and Ce Zhang. Contrastive learning for unsupervised domain adaptation of time series. In *The Eleventh International Conference on Learning Representations*, 2023.

Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.

Sanjay Purushotham, Wilka Carvalho, Tanachat Nilanon, and Yan Liu. Variational recurrent adversarial deep domain adaptation. In *International conference on learning representations*, 2017.

Sanjay Purushotham, Chuizheng Meng, Zhengping Che, and Yan Liu. Benchmarking deep learning models on large healthcare datasets. *Journal of biomedical informatics*, 83:112–134, 2018.

Peijie Qiu, Wenhui Zhu, Sayantan Kumar, Xiwen Chen, Xiaotong Sun, Jin Yang, Abolfazl Razi, Yalin Wang, and Aristeidis Sotiras. Multimodal variational autoencoder: a barycentric view. *arXiv preprint arXiv:2412.20487*, 2024.

Joaquin Quiñonero-Candela, Masashi Sugiyama, Anton Schwaighofer, and Neil D Lawrence. *Dataset shift in machine learning*. Mit Press, 2022.

Mohamed Ragab, Emadeldeen Eldele, Wee Ling Tan, Chuan-Sheng Foo, Zhenghua Chen, Min Wu, Chee-Keong Kwoh, and Xiaoli Li. Adatime: A benchmarking suite for domain adaptation on time series data. *ACM Transactions on Knowledge Discovery from Data*, 17 (8):1–18, 2023.

Pengxiang Shi, Wenwen Ye, and Zheng Qin. Self-supervised pre-training for time series classification. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2021.

Yongjie Shi, Xianghua Ying, and Jinfa Yang. Deep unsupervised domain adaptation with time series sensor data: A survey. *Sensors*, 22(15):5507, 2022.

Benjamin Shickel, Patrick James Tighe, Azra Bihorac, and Parisa Rashidi. Deep ehr: a survey of recent advances in deep learning techniques for electronic health record (ehr) analysis. *IEEE journal of biomedical and health informatics*, 22(5):1589–1604, 2017.

Baochen Sun and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. In *Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14*, pages 443–450. Springer, 2016.

Chi Ian Tang, Ignacio Perez-Pozuelo, Dimitris Spathis, and Cecilia Mascolo. Exploring contrastive learning in human activity recognition for healthcare. *arXiv preprint arXiv:2011.11542*, 2020.

Yonglong Tian, Dilip Krishnan, and Phillip Isola. Contrastive multiview coding. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16*, pages 776–794. Springer, 2020.

Sana Tonekaboni, Danny Eytan, and Anna Goldenberg. Unsupervised representation learning for time series with temporal neighborhood coding. In *International Conference on Learning Representations*, 2021.

Michael Tschannen, Josip Djolonga, Paul K. Rubenstein, Sylvain Gelly, and Mario Lucic. On mutual information maximization for representation learning. In *International Conference on Learning Representations*, 2020.

Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7167–7176, 2017.

A Vaswani. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017.

Sinong Wang, Belinda Z. Li, Madian Khabsa, Han Fang, and Hao Ma. Linformer: Self-attention with linear complexity, 2020.

Yihe Wang, Yu Han, Haishuai Wang, and Xiang Zhang. Contrast everything: A hierarchical contrastive framework for medical time-series. *Advances in Neural Information Processing Systems*, 36, 2024.

Manuel Weber, Maximilian Auch, Christoph Doblander, Peter Mandl, and Hans-Arno Jacobsen. Transfer learning with time series data: a systematic mapping study. *Ieee Access*, 9:165409–165432, 2021.

Jianfeng Wen, Nan Zhang, Xuzhe Lu, Zhongyi Hu, and Hui Huang. Mgformer: Multi-group transformer for multivariate time series classification. *Engineering Applications of Artificial Intelligence*, 133:108633, 2024.

Kristoffer Wickstrøm, Michael Kampffmeyer, Karl Øyvind Mikalsen, and Robert Jenssen. Mixing up contrastive learning: Self-supervised representation learning for time series. *Pattern Recognition Letters*, 155:54–61, 2022.

Garrett Wilson, Janardhan Rao Doppa, and Diane J Cook. Multi-source deep domain adaptation with weak supervision for time-series sensor data. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 1768–1778, 2020.

Gerald Woo, Chenghao Liu, Doyen Sahoo, Akshat Kumar, and Steven Hoi. CoST: Contrastive learning of disentangled seasonal-trend representations for time series forecasting. In *International Conference on Learning Representations*, 2022.

Feng Xie, Han Yuan, Yilin Ning, Marcus Eng Hock Ong, Mengling Feng, Wynne Hsu, Bibhas Chakraborty, and

Nan Liu. Deep learning for temporal data representation in electronic health records: A systematic review of challenges and methodologies. *Journal of biomedical informatics*, 126:103980, 2022.

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, pages 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*, pages 2114–2124, 2021.

Kexin Zhang, Qingsong Wen, Chaoli Zhang, Rongyao Cai, Ming Jin, Yong Liu, James Y Zhang, Yuxuan Liang, Guansong Pang, Dongjin Song, et al. Self-supervised learning for time series analysis: Taxonomy, progress, and prospects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

Xiang Zhang, Ziyuan Zhao, Theodoros Tsiligkaridis, and Marinka Zitnik. Self-supervised contrastive pre-training for time series via time-frequency consistency. *Advances in Neural Information Processing Systems*, 35:3988–4003, 2022.

Appendix A. Theoretical Foundations

Preliminary Definitions Let $\mathcal{D}_S$ and $\mathcal{D}_T$ denote the source and target domain distributions over the input space $\mathcal{X}$ and label space $\mathcal{Y}$. For a hypothesis $\varphi : \mathcal{X} \to \mathcal{Y}$, its classification error on a domain $\mathcal{D}$ is:

$$\epsilon_{\mathcal{D}}(\varphi) \;=\; \mathbb{E}_{(X,y)\sim\mathcal{D}}\Big[\,\mathbb{I}\big(\varphi(X) \neq y\big)\Big].$$

In unsupervised domain adaptation, we typically have labels only for one domain (or no labels if we do purely contrastive pre-training), and we aim to learn a predictor φ that performs well on the target domain $\mathcal{D}_T$. Given a sufficiently rich hypothesis class Φ (e.g., with finite VC-dimension), the classical domain-adaptation bound from Ben-David et al. (2010) states:

$$\epsilon_T(\varphi) \;\le\; \epsilon_S(\varphi) \;+\; d_\Phi\big(\mathcal{D}_S, \mathcal{D}_T\big) \;+\; \kappa,$$

where κ is the error of the ideal joint hypothesis on $\mathcal{D}_S \cup \mathcal{D}_T$, and

$$d_\Phi(\mathcal{D}_S, \mathcal{D}_T) \;=\; 2\sup_{\varphi\in\Phi}\Big|\Pr_{X\sim\mathcal{D}_S}\big[\varphi(X)=1\big] \;-\; \Pr_{X\sim\mathcal{D}_T}\big[\varphi(X)=1\big]\Big|$$

measures the distribution divergence between $\mathcal{D}_S$ and $\mathcal{D}_T$. In practice, we conduct contrastive pre-training on $\mathcal{D}_S$ to learn a domain-invariant encoder $E(\cdot)$ and feature mapping $F(\cdot)$ in an unsupervised manner.

Multi-view Contrastive Objectives We decompose each input X into three complementary views (temporal, derivative, and frequency). Our InfoNCE loss $\mathcal{L}_{CL}$ enforces that the representation of a sample is close to its positive augmentation but dissimilar from negative samples. For clarity, recall that the mutual information between two random variables A and B is defined as $I(A;B) = H(A) - H(A \mid B)$, where $H(A)$ is the entropy of A and $H(A \mid B)$ is the conditional entropy (Cover, 1999).

Proposition 1 (Multi-View Contrastive Learning and Mutual Information) *Let $\boldsymbol{z}_k$ be the feature embedding from the k-th view ($k \in \{t, d, f\}$), and suppose we have K samples in a contrastive batch. Adapting standard results (Oord et al., 2018; Tschannen et al., 2020), we obtain*

$$I\big(\boldsymbol{z}_k; \tilde{\boldsymbol{z}}_k\big) \;\ge\; \log K \;-\; \mathcal{L}_{CL}.$$

Hence, minimizing $\mathcal{L}_{CL}$ increases a lower bound on the mutual information $I(\boldsymbol{z}_k; \tilde{\boldsymbol{z}}_k)$ between $\boldsymbol{z}_k$ and $\tilde{\boldsymbol{z}}_k$. In principle, as K and the capacity of the encoder grow large, preserving high intra-view information fosters robust feature extraction for each view.

Hierarchical Multi-view Fusion Our hierarchical fusion module produces an integrated representation $\boldsymbol{H}_{\text{output}} \;=\; \big[\boldsymbol{h}_t^*,\, \boldsymbol{h}_d^*,\, \boldsymbol{h}_f^*\big], \;=\; \text{LayerNorm}(\boldsymbol{H} + \text{MHA}(\boldsymbol{H}))$ where $\boldsymbol{H} = \text{stack}(\boldsymbol{h}_t, \boldsymbol{h}_d, \boldsymbol{h}_f)$.

Proposition 2 (Inter-View Synergy via Fusion) *Under the assumption that each view $k \in \{t, d, f\}$ provides partially unique label information, the chain rule of mutual information (Han et al., 2021; Hirt et al., 2024; Oh et al., 2024a) implies:*

$$I\big(\boldsymbol{H}_{\text{output}};\, y\big) \;\ge\; \max\Big\{I(\boldsymbol{h}_t;\, y),\; I(\boldsymbol{h}_d;\, y),\; I(\boldsymbol{h}_f;\, y)\Big\}.$$

In other words, if the temporal, derivative, and frequency views are sufficiently complementary, then fusing them can capture more label-relevant structure about y than a single-view embedding alone.

Domain Adaptation under Multi-view Representations Propositions 1–2 highlight how combining the three distinct feature views (temporal, derivative, frequency) can yield more label-relevant information. Consequently, domain shifts that affect only certain views can be mitigated through the complementary nature of fusion. Suppose we have encoder $E(\cdot)$, feature extractor $F(\cdot)$, and a hierarchical feature fusion module yielding $\boldsymbol{z}_{\text{combined}} = [\boldsymbol{z}_t, \boldsymbol{z}_d, \boldsymbol{z}_f]$, along with a classifier $G(\cdot)$ on top.

Theorem 3 (Two-Stage Multi-view Adaptation) *Let Φ be a hypothesis class with finite VC-dimension, and let the source domain dataset $\mathcal{D}_S$ be sufficiently large. Assume that multi-view contrastive pre-training effectively minimizes the source embedding error $\epsilon_S(\varphi)$, producing near-optimal source representations. Furthermore, assume that each view provides complementary information about labels, thus boosting mutual information through hierarchical fusion. Then, the target-domain classification error satisfies*

$$\epsilon_T\big(G(\boldsymbol{z}_{\text{combined}})\big) \;\leq\; \epsilon_S(\varphi) \;+\; d_\Phi\big(\mathcal{D}_S, \mathcal{D}_T\big) \;+\; \kappa,$$

where κ is the minimal achievable error. Specifically, effective multi-view fusion helps reduce $d_\Phi(\mathcal{D}_S, \mathcal{D}_T)$ by leveraging complementary domain-invariant features across views, thereby controlling ϵ_T.

If the source embedding error $\epsilon_S(\varphi)$ is minimized via multi-view contrastive pre-training, the classical domain adaptation bound (Ben-David et al., 2010) indicates that $\epsilon_T(\varphi)$ is mainly influenced by $d_\Phi(\mathcal{D}_S, \mathcal{D}_T)$ and the optimal joint error κ. By learning distinct yet complementary invariances across temporal, derivative, and frequency representations, the fusion mechanism explicitly reduces effective divergence between source and target domains. Realistically, $\epsilon_S(F)$ might not be near zero if the source data are highly variable or if contrastive pre-training is imperfect. But to the extent that multi-view contrastive learning fosters robust, domain-invariant features, it reduces $\epsilon_S(F)$ and thus can help lower $\epsilon_T(F)$.

Appendix B. Experimental Details

The source code can be accessed at https://github.com/yongkyung-oh/Multi-View_Contrastive_Learning. The preprocessing steps and experimental pipeline follow the methodologies described in Zhang et al. (2022)[3] and Dong et al. (2024)[4]. For further details, please refer to their original publications.

B.1. Datasets

- SleepEEG[5] (Kemp et al., 2000): This dataset contains 153 whole-night sleep electroencephalography (EEG) recordings collected from 82 healthy subjects. Each recording is sampled at 100 Hz using a 1-lead EEG signal. The EEG signals are segmented into non-overlapping windows of size 200, each forming one sample. Each sample is labeled with one of five sleep stages: Wake (W), Non-rapid Eye Movement (N1, N2, N3), and Rapid Eye Movement (REM). This segmentation results in 371,055 samples.

- ECG[6] (Clifford et al., 2017): The 2017 PhysioNet Challenge dataset focuses on classifying single-lead electrocardiogram (ECG) recordings into four classes: normal sinus rhythm, atrial fibrillation, alternative rhythm, or noisy recordings. The signals are sampled at 300 Hz and preprocessed into 5-second samples using a fixed-length window of 1,500 observations.

- Epilepsy[7] (Andrzejak et al., 2001): This dataset consists of single-channel EEG recordings from 500 subjects, with each recording lasting 23.6 seconds and sampled at 178 Hz. The recordings are divided into 11,500 1-second samples, labeled according to five states: eyes open, eyes closed, healthy brain regions, tumor regions, and seizure episodes. For binary classification (seizure or non-seizure), the first four classes are merged. For fine-tuning, a subset of 60 samples (30 per class) is used, along with a validation set of 20 samples (10 per class). The remaining 11,420 samples are utilized for evaluation.

- FD (or FD-B)[8] (Lessmeier et al., 2016): This dataset, derived from an electromechanical drive system, monitor the condition of rolling bearings under varying operational conditions. The data include three

3. https://github.com/mims-harvard/TFC-pretraining
4. https://github.com/thuml/SimMTM
5. https://www.physionet.org/content/sleep-edfx/1.0.0/
6. https://physionet.org/content/challenge-2017/1.0.0/
7. https://timeseriesclassification.com/description.php?Dataset=Epilepsy2
8. https://mb.uni-paderborn.de/en/kat/main-research/datacenter/bearing-datacenter/data-sets-and-download

classes: undamaged, inner ring damage, and outer ring damage. Original recordings (sampled at 64k Hz for 4 seconds) are processed into samples using a sliding window of 5,120 observations with sliding windows with shift lengths of 1,024 or 4,096.

- Gesture[9] (Liu et al., 2009): This dataset comprises accelerometer measurements capturing eight distinct hand gestures, each defined by different paths of hand movement. These gestures include swiping left, right, up, and down; waving in a counterclockwise or clockwise circle; waving in a square; and waving a right arrow. The gestures are categorized into eight classes, with 440 samples in total (55 per class), as retrieved from the UCR database. The accelerometer data, likely sampled at 100 Hz, consist of three channels representing acceleration along three coordinate axes.

- EMG[10] (Goldberger et al., 2000): This dataset consists of single-channel electromyogram (EMG) recordings from the tibialis anterior muscle of three volunteers, each diagnosed with either neuropathy or myopathy. The recordings, sampled at 4k Hz, are segmented into fixed-length samples of 1,500 observations. Each volunteer (such as, disorder type) serves as a distinct classification category.

B.2. Benchmark methods

- **TST** (Zerveas et al., 2021): A Transformer-based framework for multivariate time series representation learning. It utilizes self-attention mechanisms to capture long-term dependencies and demonstrates robustness across diverse datasets.

- **TS-SD** (Shi et al., 2021): A self-supervised learning method for time series that employs sliding windows to generate input pairs for contrastive learning. It focuses on learning representations invariant to temporal distortions using DTW (Dynamic Time Warping).

- **SimCLR** (Tang et al., 2020): Adapted from the computer vision domain (Chen et al., 2020), SimCLR applies contrastive learning to time series by maximizing agreement between augmented views, enhancing representation quality.

- **TS-TCC** (Eldele et al., 2021): Learns time series representations through temporal and contextual contrasting, effectively capturing both local and global temporal dynamics.

- **CLOCS** (Kiyasseh et al., 2021): A contrastive learning framework for cardiac signals tailored to handle variations across space, time, and patients, with applications in clinical context.

- **Mixing-up** (Wickstrøm et al., 2022): Combines time series samples through interpolation to generate augmented data, improving model robustness by introducing smooth variations.

- **TS2Vec** (Yue et al., 2022): A universal time series representation framework using hierarchical contrastive loss to capture both temporal and hierarchical structures.

- **TF-C** (Zhang et al., 2022): Exploits temporal and frequency domain features for contrastive learning, producing robust, domain-invariant representations.

- **CoST** (Woo et al., 2022): Disentangles seasonal and trend representations in time series using contrastive learning, enhancing interpretability and robustness.

- **Ti-MAE** (Li et al., 2023): A masked autoencoder for time series data that reconstructs missing time points, learning robust and generalized representations.

- **SimMTM** (Dong et al., 2024): Leverages masked time points for representation learning through weighted aggregation of neighboring points outside the manifold, enabling generalization to time series.

9. http://www.timeseriesclassification.com/description.php?Dataset=UWaveGestureLibrary
10. https://physionet.org/content/emgdb/1.0.0/

B.3. Implementation Details of the Proposed Framework

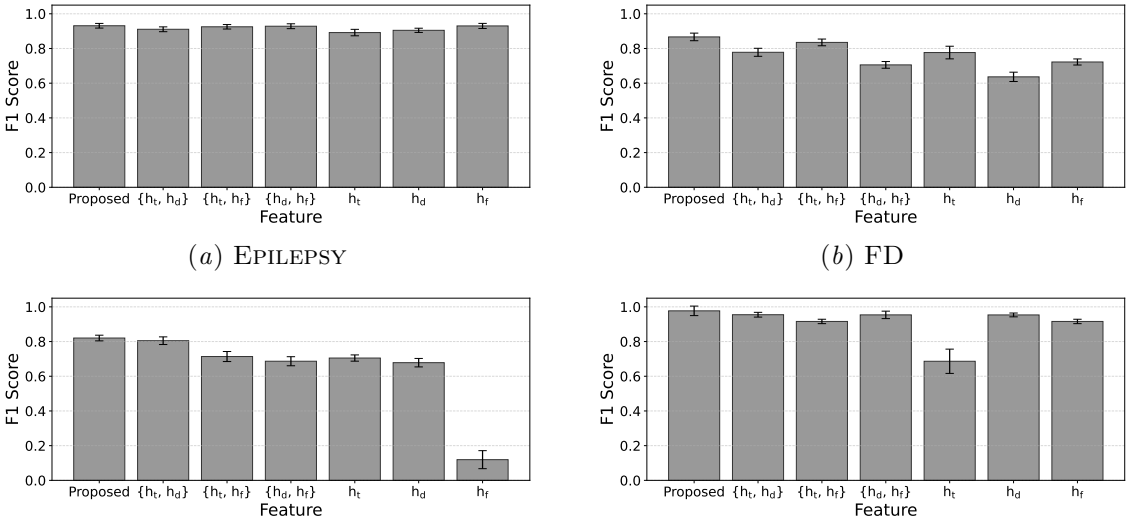
The length of each input time series is standardized to 256 using interpolation. The model architecture was kept consistent across the three features, $\{\boldsymbol{h}_t, \boldsymbol{h}_d, \boldsymbol{h}_f\}$. For $k \in \{t, d, f\}$, we implemented a transformer-based encoder E_k with a hidden size of 128, 3 layers, and 4 attention heads. The feature extractor F_k was a two-layer feedforward network with `ReLU` activation, and the classifier G consisted of a single linear layer.

To mitigate overfitting and enhance generalization, we employ several complementary strategies. First, our contrastive pre-training approach encourages the learning of robust, domain-invariant representations through data augmentations and a contrastive loss objective. Second, during fine-tuning we adopt a hybrid loss function that combines the contrastive loss with cross-entropy loss, thereby balancing unsupervised representation learning with supervised classification performance. Third, we integrate standard regularization technique, such as regularization, weight decay, adaptive learning rate scheduling, and early stopping.

Due to the differing sizes of the datasets in the pre-training and fine-tuning stages, we used distinct batch sizes and epochs for each phase. For pre-training, a batch size of 128 and 200 epochs were used, while fine-tuning employed a batch size of 16 and 100 epochs. In both cases, the learning rate was set to 10^{-3} with a weight decay of 10^{-5}. To optimize training efficiency and adaptively adjust the learning rate, we employed a learning rate scheduler. Specifically, we used a `ReduceLROnPlateau` scheduler, which monitors the validation loss and reduces the learning rate by a factor of 0.1 if the validation loss does not improve for 10 consecutive epochs. Additionally, we incorporated early stopping to terminate training if the validation loss did not improve for 20 consecutive epochs, further enhancing computational efficiency.

Appendix C. Detailed Results

We present our experimental findings in Figure 4, which compares the F1 scores of different feature combinations pre-trained on the SleepEEG dataset across four target domains: Epilepsy, FD, Gesture, and EMG. Our analysis reveals that the Proposed model consistently achieves the highest F1 score across all tasks, demonstrating the effectiveness of our hierarchical fusion strategy. These results validate the robustness of our hierarchical fusion approach, which provides more generalizable representations across diverse tasks. However, the varying sensitivity of different tasks to individual feature combinations underscores the importance of domain expertise in optimal feature selection.



(a) Epilepsy (b) FD

(c) Gesture (d) EMG

Figure 4: F1 score comparison on feature combinations of our method pre-trained on SleepEEG dataset

Table 7: Performance comparison with benchmark method: SleepEEG → FD
(mean ± standard deviation over five runs; 0.000 indicates unreported variance in the referenced studies)

Method \ Metric	Accuracy	Precision	Recall	F1 score
TST (Zerveas et al., 2021)	$0.464_{\pm 0.000}$	$0.416_{\pm 0.000}$	$0.455_{\pm 0.000}$	$0.413_{\pm 0.000}$
TS-SD (Shi et al., 2021)	$0.557_{\pm 0.021}$	$0.571_{\pm 0.054}$	$0.605_{\pm 0.027}$	$0.570_{\pm 0.033}$
SimCLR (Tang et al., 2020)	$0.492_{\pm 0.044}$	$0.545_{\pm 0.102}$	$0.476_{\pm 0.089}$	$0.422_{\pm 0.114}$
TS-TCC (Eldele et al., 2021)	$0.550_{\pm 0.022}$	$0.528_{\pm 0.029}$	$0.640_{\pm 0.018}$	$0.542_{\pm 0.034}$
CLOCS (Kiyasseh et al., 2021)	$0.493_{\pm 0.031}$	$0.482_{\pm 0.032}$	$0.587_{\pm 0.039}$	$0.475_{\pm 0.049}$
Mixing-up (Wickstrøm et al., 2022)	$0.679_{\pm 0.025}$	$0.715_{\pm 0.034}$	$0.761_{\pm 0.020}$	$0.727_{\pm 0.023}$
TS2Vec (Yue et al., 2022)	$0.479_{\pm 0.011}$	$0.434_{\pm 0.009}$	$0.484_{\pm 0.020}$	$0.439_{\pm 0.011}$
TF-C (Zhang et al., 2022)	$0.694_{\pm 0.023}$	$0.756_{\pm 0.035}$	$0.720_{\pm 0.026}$	$0.749_{\pm 0.027}$
CoST (Woo et al., 2022)	$0.471_{\pm 0.000}$	$0.388_{\pm 0.000}$	$0.384_{\pm 0.000}$	$0.348_{\pm 0.000}$
Ti-MAE (Li et al., 2023)	$0.609_{\pm 0.000}$	$0.670_{\pm 0.000}$	$0.689_{\pm 0.000}$	$0.666_{\pm 0.000}$
SimMTM (Dong et al., 2024)	$0.694_{\pm 0.000}$	$0.742_{\pm 0.000}$	$0.764_{\pm 0.000}$	$0.751_{\pm 0.000}$
Proposed method	$\mathbf{0.879}_{\pm 0.010}$	$\mathbf{0.853}_{\pm 0.014}$	$\mathbf{0.891}_{\pm 0.009}$	$\mathbf{0.867}_{\pm 0.012}$

Table 8: Sensitivity analysis on FD with learning strategies, including different source datasets, self-supervised learning, and random initialization (For each experimental scenario, classifier $G(\cdot)$ was initialized and subsequently trained by target dataset. 'freeze scenario' indicates that $E_k(\cdot)$ and $F_k(\cdot)$ are fixed after pre-training, while only $G(\cdot)$ are optimized during fine-tuning, where $k \in \{t, d, f\}$.)

Learning Strategy \ Metric	Accuracy	Precision	Recall	F1 score
Source: SleepEEG	$0.879_{\pm 0.010}$	$0.853_{\pm 0.014}$	$0.891_{\pm 0.009}$	$0.867_{\pm 0.012}$
– freeze $E_k(\cdot)$ and $F_k(\cdot)$	$0.827_{\pm 0.013}$	$0.809_{\pm 0.011}$	$0.865_{\pm 0.009}$	$0.824_{\pm 0.006}$
Source: ECG	$0.888_{\pm 0.008}$	$0.862_{\pm 0.009}$	$0.904_{\pm 0.005}$	$0.878_{\pm 0.006}$
– freeze $E_k(\cdot)$ and $F_k(\cdot)$	$0.872_{\pm 0.009}$	$0.849_{\pm 0.014}$	$0.893_{\pm 0.008}$	$0.861_{\pm 0.008}$
Source: FD	$0.877_{\pm 0.008}$	$0.848_{\pm 0.014}$	$0.889_{\pm 0.009}$	$0.865_{\pm 0.008}$
– freeze $E_k(\cdot)$ and $F_k(\cdot)$	$0.869_{\pm 0.007}$	$0.823_{\pm 0.013}$	$0.884_{\pm 0.005}$	$0.844_{\pm 0.011}$
Random initialization	$0.882_{\pm 0.014}$	$0.866_{\pm 0.019}$	$0.900_{\pm 0.015}$	$0.877_{\pm 0.006}$

Table 9: Comprehensive analysis of the proposed framework's components on FD
(1) hierarchical feature fusion, (2) fine-tuning loss configurations with λ, and (3) combinations of features $\{\boldsymbol{h}_t, \boldsymbol{h}_d, \boldsymbol{h}_f\}$. (The proposed method utilizes all three features with $\lambda = 0.1$ as the default configuration.)

Model Component \ Metric	Accuracy	Precision	Recall	F1 score
Proposed method	$0.879_{\pm 0.010}$	$0.853_{\pm 0.014}$	$0.891_{\pm 0.009}$	$0.867_{\pm 0.012}$
– use $\mathcal{L}_{CE}$ only	$0.879_{\pm 0.010}$	$0.853_{\pm 0.014}$	$0.890_{\pm 0.010}$	$0.867_{\pm 0.012}$
w/o feature fusion	$0.856_{\pm 0.019}$	$0.838_{\pm 0.020}$	$0.877_{\pm 0.015}$	$0.851_{\pm 0.012}$
– use $\mathcal{L}_{CE}$ only	$0.847_{\pm 0.027}$	$0.836_{\pm 0.022}$	$0.867_{\pm 0.021}$	$0.845_{\pm 0.017}$
$\{\boldsymbol{h}_t, \boldsymbol{h}_d\}$	$0.765_{\pm 0.006}$	$0.765_{\pm 0.018}$	$0.817_{\pm 0.011}$	$0.778_{\pm 0.013}$
$\{\boldsymbol{h}_t, \boldsymbol{h}_f\}$	$0.854_{\pm 0.008}$	$0.818_{\pm 0.018}$	$0.878_{\pm 0.011}$	$0.835_{\pm 0.009}$
$\{\boldsymbol{h}_d, \boldsymbol{h}_f\}$	$0.740_{\pm 0.014}$	$0.688_{\pm 0.011}$	$0.769_{\pm 0.009}$	$0.705_{\pm 0.009}$
$\boldsymbol{h}_t$ only	$0.781_{\pm 0.033}$	$0.755_{\pm 0.029}$	$0.824_{\pm 0.020}$	$0.777_{\pm 0.026}$
$\boldsymbol{h}_d$ only	$0.671_{\pm 0.015}$	$0.625_{\pm 0.011}$	$0.705_{\pm 0.011}$	$0.637_{\pm 0.017}$
$\boldsymbol{h}_f$ only	$0.768_{\pm 0.006}$	$0.711_{\pm 0.008}$	$0.757_{\pm 0.023}$	$0.722_{\pm 0.008}$

Table 10: Performance comparison with benchmark methods: SleepEEG → Gesture (mean ± standard deviation over five runs; 0.000 indicates unreported variance in the referenced studies)

Method \ Metric	Accuracy	Precision	Recall	F1 score
TST (Zerveas et al., 2021)	$0.692_{\pm 0.000}$	$0.666_{\pm 0.000}$	$0.692_{\pm 0.000}$	$0.660_{\pm 0.000}$
TS-SD (Shi et al., 2021)	$0.692_{\pm 0.044}$	$0.670_{\pm 0.047}$	$0.687_{\pm 0.049}$	$0.666_{\pm 0.044}$
SimCLR (Tang et al., 2020)	$0.480_{\pm 0.059}$	$0.595_{\pm 0.162}$	$0.541_{\pm 0.195}$	$0.496_{\pm 0.187}$
TS-TCC (Eldele et al., 2021)	$0.719_{\pm 0.035}$	$0.714_{\pm 0.035}$	$0.717_{\pm 0.037}$	$0.698_{\pm 0.036}$
CLOCS (Kiyasseh et al., 2021)	$0.443_{\pm 0.052}$	$0.424_{\pm 0.079}$	$0.443_{\pm 0.052}$	$0.401_{\pm 0.060}$
Mixing-up (Wickstrøm et al., 2022)	$0.693_{\pm 0.023}$	$0.672_{\pm 0.023}$	$0.693_{\pm 0.023}$	$0.650_{\pm 0.031}$
TS2Vec (Yue et al., 2022)	$0.692_{\pm 0.033}$	$0.655_{\pm 0.036}$	$0.685_{\pm 0.035}$	$0.657_{\pm 0.039}$
TF-C (Zhang et al., 2022)	$0.764_{\pm 0.020}$	$0.773_{\pm 0.036}$	$0.743_{\pm 0.027}$	$0.757_{\pm 0.031}$
CoST (Woo et al., 2022)	$0.683_{\pm 0.000}$	$0.653_{\pm 0.000}$	$0.683_{\pm 0.000}$	$0.664_{\pm 0.000}$
Ti-MAE (Li et al., 2023)	$0.719_{\pm 0.000}$	$0.704_{\pm 0.000}$	$0.768_{\pm 0.000}$	$0.684_{\pm 0.000}$
SimMTM (Dong et al., 2024)	$0.800_{\pm 0.000}$	$0.790_{\pm 0.000}$	$0.800_{\pm 0.000}$	$0.787_{\pm 0.000}$
Proposed method	$\mathbf{0.832}_{\pm 0.010}$	$\mathbf{0.830}_{\pm 0.010}$	$\mathbf{0.832}_{\pm 0.010}$	$\mathbf{0.820}_{\pm 0.006}$

Table 11: Sensitivity analysis on Gesture with learning strategies, including different source datasets, self-supervised learning, and random initialization (For each experimental scenario, classifier $G(\cdot)$ was initialized and subsequently trained by target dataset. 'freeze scenario' indicates that $E_k(\cdot)$ and $F_k(\cdot)$ are fixed after pre-training, while only $G(\cdot)$ are optimized during fine-tuning, where $k \in \{t, d, f\}$.)

Learning Strategy \ Metric	Accuracy	Precision	Recall	F1 score
Source: SleepEEG	$0.832_{\pm 0.010}$	$0.830_{\pm 0.010}$	$0.832_{\pm 0.010}$	$0.820_{\pm 0.006}$
– freeze $E_k(\cdot)$ and $F_k(\cdot)$	$0.833_{\pm 0.007}$	$0.832_{\pm 0.008}$	$0.833_{\pm 0.007}$	$0.820_{\pm 0.006}$
Source: ECG	$0.817_{\pm 0.011}$	$0.815_{\pm 0.019}$	$0.817_{\pm 0.011}$	$0.808_{\pm 0.012}$
– freeze $E_k(\cdot)$ and $F_k(\cdot)$	$0.735_{\pm 0.008}$	$0.740_{\pm 0.006}$	$0.735_{\pm 0.008}$	$0.719_{\pm 0.008}$
Source: Gesture	$0.792_{\pm 0.007}$	$0.788_{\pm 0.008}$	$0.792_{\pm 0.007}$	$0.786_{\pm 0.006}$
– freeze $E_k(\cdot)$ and $F_k(\cdot)$	$0.792_{\pm 0.007}$	$0.788_{\pm 0.008}$	$0.792_{\pm 0.007}$	$0.786_{\pm 0.006}$
Random initialization	$0.790_{\pm 0.010}$	$0.779_{\pm 0.015}$	$0.790_{\pm 0.010}$	$0.773_{\pm 0.008}$

Table 12: Comprehensive analysis of the proposed framework's components on Gesture (1) hierarchical feature fusion, (2) fine-tuning loss configurations with λ, and (3) combinations of features $\{\boldsymbol{h}_t, \boldsymbol{h}_d, \boldsymbol{h}_f\}$. (The proposed method utilizes all three features with $\lambda = 0.1$ as the default configuration.)

Model Component \ Metric	Accuracy	Precision	Recall	F1 score
Proposed method	$0.832_{\pm 0.010}$	$0.830_{\pm 0.010}$	$0.832_{\pm 0.010}$	$0.820_{\pm 0.006}$
– use $\mathcal{L}_{CE}$ only	$0.833_{\pm 0.007}$	$0.832_{\pm 0.008}$	$0.833_{\pm 0.007}$	$0.820_{\pm 0.006}$
w/o feature fusion	$0.808_{\pm 0.007}$	$0.807_{\pm 0.006}$	$0.808_{\pm 0.007}$	$0.798_{\pm 0.011}$
– use $\mathcal{L}_{CE}$ only	$0.798_{\pm 0.012}$	$0.791_{\pm 0.016}$	$0.798_{\pm 0.012}$	$0.785_{\pm 0.012}$
$\{\boldsymbol{h}_t, \boldsymbol{h}_d\}$	$0.815_{\pm 0.012}$	$0.813_{\pm 0.008}$	$0.815_{\pm 0.012}$	$0.805_{\pm 0.012}$
$\{\boldsymbol{h}_t, \boldsymbol{h}_f\}$	$0.743_{\pm 0.016}$	$0.729_{\pm 0.034}$	$0.743_{\pm 0.016}$	$0.714_{\pm 0.019}$
$\{\boldsymbol{h}_d, \boldsymbol{h}_f\}$	$0.698_{\pm 0.019}$	$0.711_{\pm 0.024}$	$0.698_{\pm 0.019}$	$0.687_{\pm 0.016}$
$\boldsymbol{h}_t$ only	$0.732_{\pm 0.012}$	$0.726_{\pm 0.032}$	$0.732_{\pm 0.012}$	$0.705_{\pm 0.008}$
$\boldsymbol{h}_d$ only	$0.690_{\pm 0.016}$	$0.697_{\pm 0.025}$	$0.690_{\pm 0.016}$	$0.678_{\pm 0.014}$
$\boldsymbol{h}_f$ only	$0.202_{\pm 0.033}$	$0.122_{\pm 0.066}$	$0.202_{\pm 0.033}$	$0.119_{\pm 0.042}$

Table 13: Performance comparison with benchmark methods: SleepEEG → EMG
(mean ± standard deviation over five runs; 0.000 indicates unreported variance in the referenced studies)

Method \ Metric	Accuracy	Precision	Recall	F1 score
TST (Zerveas et al., 2021)	$0.783_{\pm 0.000}$	$0.771_{\pm 0.000}$	$0.803_{\pm 0.000}$	$0.689_{\pm 0.000}$
TS-SD (Shi et al., 2021)	$0.461_{\pm 0.000}$	$0.155_{\pm 0.000}$	$0.333_{\pm 0.000}$	$0.211_{\pm 0.000}$
SimCLR (Tang et al., 2020)	$0.615_{\pm 0.058}$	$0.516_{\pm 0.172}$	$0.499_{\pm 0.121}$	$0.471_{\pm 0.149}$
TS-TCC (Eldele et al., 2021)	$0.789_{\pm 0.019}$	$0.585_{\pm 0.097}$	$0.631_{\pm 0.099}$	$0.590_{\pm 0.095}$
CLOCS (Kiyasseh et al., 2021)	$0.699_{\pm 0.032}$	$0.511_{\pm 0.075}$	$0.515_{\pm 0.029}$	$0.514_{\pm 0.041}$
Mixing-up (Wickstrøm et al., 2022)	$0.302_{\pm 0.053}$	$0.110_{\pm 0.013}$	$0.258_{\pm 0.046}$	$0.154_{\pm 0.020}$
TS2Vec (Yue et al., 2022)	$0.785_{\pm 0.032}$	$0.804_{\pm 0.075}$	$0.679_{\pm 0.040}$	$0.677_{\pm 0.050}$
TF-C (Zhang et al., 2022)	$0.817_{\pm 0.029}$	$0.727_{\pm 0.035}$	$0.816_{\pm 0.029}$	$0.768_{\pm 0.031}$
CoST (Woo et al., 2022)	$0.517_{\pm 0.000}$	$0.491_{\pm 0.000}$	$0.421_{\pm 0.000}$	$0.353_{\pm 0.000}$
Ti-MAE (Li et al., 2023)	$0.700_{\pm 0.000}$	$0.703_{\pm 0.000}$	$0.634_{\pm 0.000}$	$0.709_{\pm 0.000}$
SimMTM (Dong et al., 2024)	$0.976_{\pm 0.000}$	$\mathbf{0.983_{\pm 0.000}}$	$0.980_{\pm 0.000}$	$\mathbf{0.981_{\pm 0.000}}$
Proposed method	$\mathbf{0.980_{\pm 0.012}}$	$0.976_{\pm 0.022}$	$\mathbf{0.984_{\pm 0.010}}$	$0.977_{\pm 0.017}$

Table 14: Sensitivity analysis on EMG with learning strategies, including different source datasets, self-supervised learning, and random initialization (For each experimental scenario, classifier $G(\cdot)$ was initialized and subsequently trained by target dataset. 'freeze scenario' indicates that $E_k(\cdot)$ and $F_k(\cdot)$ are fixed after pre-training, while only $G(\cdot)$ are optimized during fine-tuning, where $k \in \{t, d, f\}$.)

Learning Strategy \ Metric	Accuracy	Precision	Recall	F1 score
Source: SleepEEG	$0.980_{\pm 0.012}$	$0.976_{\pm 0.022}$	$0.984_{\pm 0.010}$	$0.977_{\pm 0.017}$
– freeze $E_k(\cdot)$ and $F_k(\cdot)$	$0.971_{\pm 0.012}$	$0.973_{\pm 0.018}$	$0.976_{\pm 0.010}$	$0.973_{\pm 0.012}$
Source: ECG	$0.995_{\pm 0.012}$	$0.997_{\pm 0.008}$	$0.996_{\pm 0.010}$	$0.992_{\pm 0.019}$
– freeze $E_k(\cdot)$ and $F_k(\cdot)$	$0.995_{\pm 0.012}$	$0.997_{\pm 0.008}$	$0.992_{\pm 0.019}$	$0.991_{\pm 0.021}$
Source: EMG	$0.976_{\pm 0.001}$	$0.974_{\pm 0.018}$	$0.980_{\pm 0.001}$	$0.964_{\pm 0.011}$
– freeze $E_k(\cdot)$ and $F_k(\cdot)$	$0.976_{\pm 0.001}$	$0.967_{\pm 0.021}$	$0.971_{\pm 0.022}$	$0.958_{\pm 0.004}$
Random initialization	$0.980_{\pm 0.012}$	$0.978_{\pm 0.021}$	$0.984_{\pm 0.010}$	$0.972_{\pm 0.019}$

Table 15: Comprehensive analysis of the proposed framework's components on EMG
(1) hierarchical feature fusion, (2) fine-tuning loss configurations with λ, and (3) combinations of features $\{\boldsymbol{h}_t, \boldsymbol{h}_d, \boldsymbol{h}_f\}$. (The proposed method utilizes all three features with $\lambda = 0.1$ as the default configuration.)

Model Component \ Metric	Accuracy	Precision	Recall	F1 score
Proposed method	$0.980_{\pm 0.012}$	$0.976_{\pm 0.022}$	$0.984_{\pm 0.010}$	$0.977_{\pm 0.017}$
– use $\mathcal{L}_{CE}$ only	$0.976_{\pm 0.001}$	$0.973_{\pm 0.018}$	$0.980_{\pm 0.001}$	$0.973_{\pm 0.012}$
w/o feature fusion	$0.971_{\pm 0.012}$	$0.973_{\pm 0.018}$	$0.976_{\pm 0.010}$	$0.973_{\pm 0.012}$
– use $\mathcal{L}_{CE}$ only	$0.966_{\pm 0.014}$	$0.969_{\pm 0.025}$	$0.973_{\pm 0.012}$	$0.969_{\pm 0.019}$
$\{\boldsymbol{h}_t, \boldsymbol{h}_d\}$	$0.976_{\pm 0.001}$	$0.981_{\pm 0.001}$	$0.943_{\pm 0.022}$	$0.955_{\pm 0.004}$
$\{\boldsymbol{h}_t, \boldsymbol{h}_f\}$	$0.941_{\pm 0.014}$	$0.929_{\pm 0.018}$	$0.912_{\pm 0.010}$	$0.916_{\pm 0.003}$
$\{\boldsymbol{h}_d, \boldsymbol{h}_f\}$	$0.971_{\pm 0.012}$	$0.956_{\pm 0.019}$	$0.958_{\pm 0.032}$	$0.954_{\pm 0.011}$
$\boldsymbol{h}_t$ only	$0.873_{\pm 0.012}$	$0.754_{\pm 0.163}$	$0.698_{\pm 0.034}$	$0.686_{\pm 0.060}$
$\boldsymbol{h}_d$ only	$0.976_{\pm 0.001}$	$0.981_{\pm 0.001}$	$0.933_{\pm 0.001}$	$0.953_{\pm 0.001}$
$\boldsymbol{h}_f$ only	$0.941_{\pm 0.014}$	$0.929_{\pm 0.018}$	$0.907_{\pm 0.011}$	$0.916_{\pm 0.003}$

FD Table 7 shows the proposed method outperforming benchmarks on the FD dataset, showcasing its adaptability to industrial time series. Table 8 highlights the robustness of learning strategies, and Table 9 confirms the significant impact of multi-view fusion over single features.

Gesture Table 10 demonstrates state-of-the-art performance by the proposed method on the Gesture dataset. Table 11 validates SleepEEG as the best pre-training source, while Table 12 emphasizes the importance of hierarchical feature fusion.

EMG Table 13 highlights the superior performance of the proposed method on the EMG dataset. Table 14 confirms the value of fine-tuning on pre-trained models, and Table 15 showcase the importance of each feature. Notably, the self-supervised learning approach achieves superior performance compared to other.

It is important to note that the optimal source dataset varies depending on the target domain, highlighting the significance of incorporating domain knowledge and sophisticated feature selection approach.

Appendix D. Comprehensive Practical Analysis

D.1. Practical Computation Strategy

Table 16 shows that half-precision arithmetic by AMP (Micikevicius et al., 2018) attains competitive classification metrics across different pre-training sources on Epilepsy. Moreover, we observed an average batch inference time of 0.0519 ± 0.0002 s with AMP versus 0.0823 ± 0.0003 s in full precision mode, indicating a tangible speedup. Note that these computation times may vary depending on hardware specifications, batch sizes, and other implementation details. Even so, AMP can reduce memory usage and computational cost during training while also accelerating inference, helping to offset the additional overhead associated with multi-view feature extraction and still maintaining robust domain adaptation performance.

Table 16: Comparative analysis of performance on Epilepsy using three different source datasets (SleepEEG, and ECG) under automatic mixed precision (AMP) vs. full-precision settings.

Setting	Domain Adaptation	Accuracy	Precision	Recall	F1 score
with AMP	SleepEEG → Epilepsy	$0.958_{\pm 0.002}$	$0.936_{\pm 0.004}$	$0.935_{\pm 0.004}$	$0.931_{\pm 0.003}$
	ECG → Epilepsy	$0.956_{\pm 0.002}$	$0.934_{\pm 0.004}$	$0.940_{\pm 0.008}$	$0.927_{\pm 0.003}$
full-precision	SleepEEG → Epilepsy	$0.957_{\pm 0.003}$	$0.939_{\pm 0.014}$	$0.942_{\pm 0.004}$	$0.932_{\pm 0.004}$
	ECG → Epilepsy	$0.955_{\pm 0.004}$	$0.934_{\pm 0.010}$	$0.944_{\pm 0.005}$	$0.928_{\pm 0.005}$

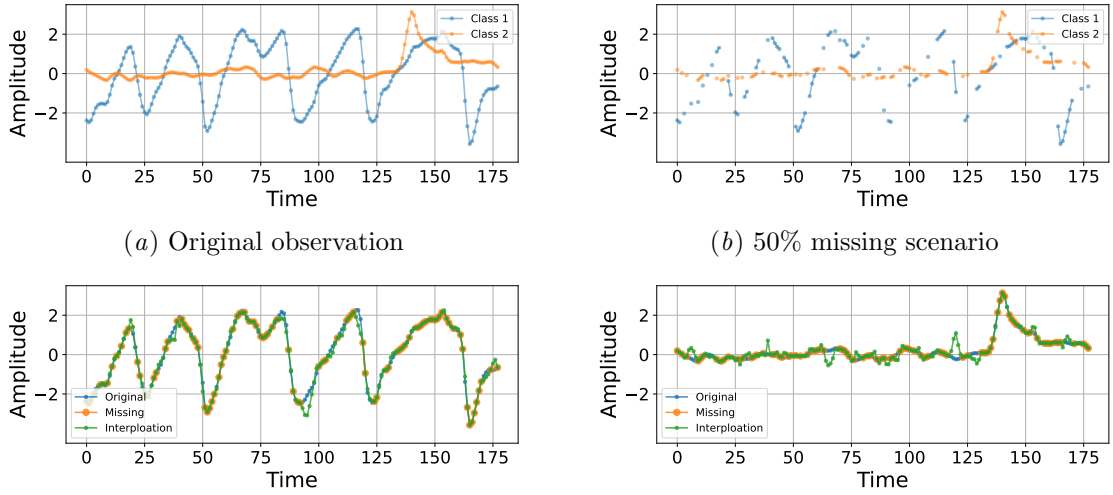
D.2. Handling Irregularly-Sampled Time Series Data

Real-world medical or sensor data are frequently characterized by irregular sampling intervals and missing values. A straightforward way to address this in our framework is to use the same interpolation mechanism adopted for derivative computation. Figure 5 (a) shows the original time series, (b) illustrates an artificially introduced missing scenario where we randomly remove about 50% of observations from each sample, independently, leading to irregular sampling. Figure 5 (c) and (d) depict the spline-based interpolation used to recover a uniform grid. Although interpolation closely aligns with the original data, unavoidable deviations from missing observations introduce a dataset shift.

Furthermore, Figure 6 compares the extracted features (temporal, derivative, frequency) obtained from the original versus interpolated signals from the selected samples. The black curves indicate features from the original data, while the red curves show those derived via interpolation-based resampling. In practice, this interpolation preserves the core patterns in all three feature views.

Table 17 reports classification for two scenarios: using the original data, and using an interpolation-based method to handle irregular or missing observations. We pre-trained the model on regularly sampled time series but performed fine-tuning with irregularly-sampled data. Despite a slight performance drop in the

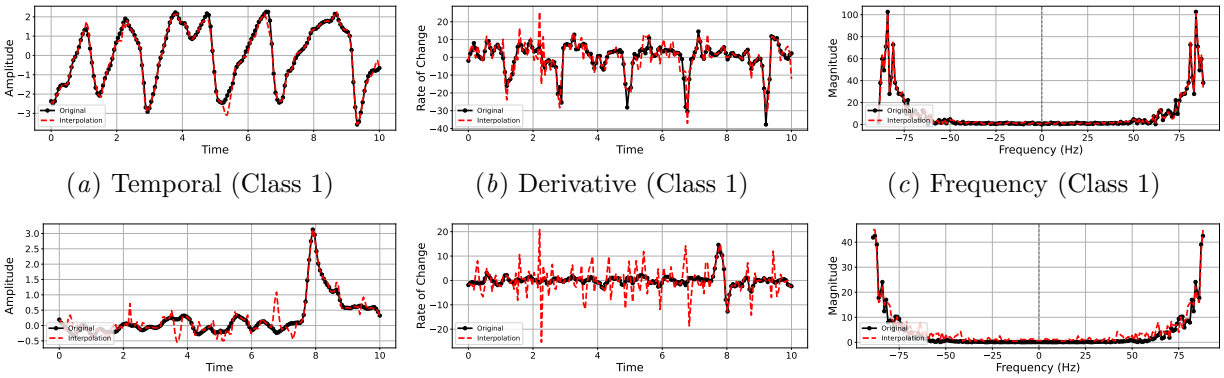
interpolated setting, the overall metrics remain feasible, indicating that our framework is robust even with irregularly sampling or missing data. In future work, we intend to explore more sophisticated methods for managing irregularly-sampled time series (e.g., continuous-time models or neural differential equations) and further enrich our framework's ability to handle incomplete real-world data.



(*a*) Original observation (*b*) 50% missing scenario

(*c*) Interpolation (Class 1) (*d*) Interpolation (Class 2)

Figure 5: Example of handling irregularly-sampled observations in the Epilepsy dataset



(*a*) Temporal (Class 1) (*b*) Derivative (Class 1) (*c*) Frequency (Class 1)

(*d*) Temporal (Class 2) (*e*) Derivative (Class 2) (*f*) Frequency (Class 2)

Figure 6: Comparison of features computed on the original versus interpolated Epilepsy dataset

Table 17: Comparative analysis of performance on Epilepsy with irregularly-sampled scenario

Setting	Domain Adaptation	Accuracy	Precision	Recall	F1 score
Original	SleepEEG → Epilepsy	$0.958_{\pm 0.002}$	$0.936_{\pm 0.004}$	$0.935_{\pm 0.004}$	$0.931_{\pm 0.003}$
	ECG → Epilepsy	$0.956_{\pm 0.002}$	$0.934_{\pm 0.004}$	$0.940_{\pm 0.008}$	$0.927_{\pm 0.003}$
Interpolation	SleepEEG → Epilepsy	$0.951_{\pm 0.003}$	$0.937_{\pm 0.013}$	$0.929_{\pm 0.005}$	$0.922_{\pm 0.003}$
	ECG → Epilepsy	$0.951_{\pm 0.003}$	$0.927_{\pm 0.007}$	$0.931_{\pm 0.004}$	$0.923_{\pm 0.004}$

D.3. Impact of Source Dataset in Domain Adaptation

Incomplete Source Dataset Table 18 contrasts performance on Epilepsy when using the entire source dataset (100%) versus a much smaller portion (1%). Notably, even 1% of SleepEEG or ECG is larger than the full training set of Epilepsy, indicating that a relatively small fraction of a homogeneous source domain can still provide meaningful pre-training benefits. However, the larger standard deviations in the 1% setting suggest that while small-scale pre-training can be beneficial, it may not always yield stable results—indicating a trade-off between data efficiency and robustness

Table 18: Comparative analysis of performance on Epilepsy with partial source dataset

Setting	Domain Adaptation	Accuracy	Precision	Recall	F1 score
100%	SleepEEG → Epilepsy	$0.958_{\pm 0.002}$	$0.936_{\pm 0.004}$	$0.935_{\pm 0.004}$	$0.931_{\pm 0.003}$
	ECG → Epilepsy	$0.956_{\pm 0.002}$	$0.934_{\pm 0.004}$	$0.940_{\pm 0.008}$	$0.927_{\pm 0.003}$
1%	SleepEEG → Epilepsy	$0.954_{\pm 0.002}$	$0.923_{\pm 0.006}$	$0.945_{\pm 0.004}$	$0.928_{\pm 0.002}$
	ECG → Epilepsy	$0.943_{\pm 0.009}$	$0.908_{\pm 0.019}$	$0.937_{\pm 0.004}$	$0.913_{\pm 0.012}$

Catastrophic Forgetting and Overfitting While adapting a model from one domain to another, two primary risks arise: *catastrophic forgetting* of previously learned representations and *overfitting* to the new domain's limited data. In our cross-domain adaptation scenario, we mitigate these issues through a balanced loss objective that preserves domain-invariant features (via contrastive learning) while fine-tuning on the target domain (via cross-entropy loss). This joint learning scheme discourages drastic parameter updates that could wipe out essential knowledge from the source, and it helps prevent overfitting by maintaining a more generalizable feature space.

Cross-Domain vs. Multi-Source Pre-Training One possible approach to domain adaptation in time series is to merge data from multiple sources into a single, large pre-training set. Our one-to-one and one-to-many setups instead assume a relatively homogeneous source domain, making domain-invariant pattern learning more tractable and interpretable. However, when data originate from vastly different domains (e.g., EEG signals vs. industrial sensor logs), a many-to-one strategy may result in excessive heterogeneity that obscures useful temporal patterns. Developing robust solutions for large-scale, highly heterogeneous time series remains an open research direction, and we hope our systematic cross-domain experiments will serve as a foundation for future work aiming to unify multi-sourced data.

D.4. Benchmarking Across Multivariate Time Series Classification Tasks

We further evaluated our framework on ten multivariate time series classification (MTSC) datasets from the University of East Anglia (UEA) repository[11](Bagnall et al., 2018). For comparison, we use the reported performances of TST (Zerveas et al., 2021), TS-TCC (Eldele et al., 2021), TNC (Tonekaboni et al., 2021), T-Loss (Franceschi et al., 2019), TS2Vec (Yue et al., 2022), MICOS (Hao et al., 2023), and Mgformer (Wen et al., 2024) as presented in Wen et al. (2024). We train our own model purely in a self-supervised manner, without using any additional external data (such as SleepEEG or ECG) or hyperparameter tuning process.

Table 19 highlights our approach's performance on ten public multivariate time series datasets. The columns list classification accuracy for each baseline method and our proposed framework. Notably, the proposed method achieves competitive performance on a majority of the benchmarks. Table 20 examines the contribution of each architectural and training component in our proposed framework. By comparing these settings, we observe how multi-view fusion and the combined loss objective each drive consistent improvements across datasets, further validating the holistic design of our approach. Observing consistent gains when all three are combined supports our claim that integrating multiple feature views captures richer representations and yields stronger results across a wide range of time series applications.

11. https://www.timeseriesclassification.com/

Table 19: Comparative analysis of performance on MTSC

Dataset \Method	TST	TS-TCC	TNC	T-Loss	TS2Vec	MICOS	Mgformer	Proposed
AtrialFibrillation	0.067	0.267	0.133	0.133	0.200	0.333	0.400	$0.427_{\pm 0.077}$
BasicMotions	0.975	1.000	0.975	1.000	0.975	1.000	1.000	$0.995_{\pm 0.012}$
FingerMovements	0.560	0.460	0.470	0.580	0.480	0.570	0.620	$0.574_{\pm 0.027}$
Heartbeat	0.746	0.751	0.746	0.741	0.683	0.766	0.790	$0.755_{\pm 0.005}$
MotorImagery	0.500	0.610	0.500	0.580	0.510	0.500	0.530	$0.576_{\pm 0.010}$
RacketSports	0.809	0.816	0.776	0.855	0.855	0.941	0.908	$0.821_{\pm 0.019}$
SelfRegulationSCP1	0.754	0.823	0.799	0.843	0.812	0.799	0.890	$0.886_{\pm 0.010}$
SelfRegulationSCP2	0.550	0.533	0.550	0.539	0.578	0.578	0.533	$0.594_{\pm 0.017}$
StandWalkJump	0.267	0.333	0.400	0.333	0.467	0.533	0.600	$0.493_{\pm 0.077}$
UWaveGestureLibrary	0.575	0.753	0.759	0.875	0.906	0.891	0.893	$0.827_{\pm 0.035}$

Table 20: Comprehensive analysis of the proposed framework's components on MTSC

Dataset \ Model Component	Proposed method		w/o feature fusion	
	$\mathcal{L}_{CL} + \mathcal{L}_{CE}$	use $\mathcal{L}_{CE}$ only	$\mathcal{L}_{CL} + \mathcal{L}_{CE}$	use $\mathcal{L}_{CE}$ only
AtrialFibrillation	$0.427_{\pm 0.077}$	$0.413_{\pm 0.088}$	$0.373_{\pm 0.102}$	$0.360_{\pm 0.113}$
BasicMotions	$0.995_{\pm 0.012}$	$0.995_{\pm 0.012}$	$0.995_{\pm 0.012}$	$0.995_{\pm 0.012}$
FingerMovements	$0.574_{\pm 0.027}$	$0.564_{\pm 0.018}$	$0.566_{\pm 0.020}$	$0.564_{\pm 0.018}$
Heartbeat	$0.755_{\pm 0.005}$	$0.753_{\pm 0.004}$	$0.745_{\pm 0.006}$	$0.745_{\pm 0.006}$
MotorImagery	$0.576_{\pm 0.010}$	$0.572_{\pm 0.016}$	$0.562_{\pm 0.017}$	$0.548_{\pm 0.029}$
RacketSports	$0.821_{\pm 0.019}$	$0.817_{\pm 0.017}$	$0.816_{\pm 0.019}$	$0.816_{\pm 0.019}$
SelfRegulationSCP1	$0.886_{\pm 0.010}$	$0.878_{\pm 0.021}$	$0.876_{\pm 0.010}$	$0.868_{\pm 0.019}$
SelfRegulationSCP2	$0.594_{\pm 0.017}$	$0.576_{\pm 0.015}$	$0.586_{\pm 0.019}$	$0.568_{\pm 0.021}$
StandWalkJump	$0.493_{\pm 0.077}$	$0.440_{\pm 0.061}$	$0.427_{\pm 0.077}$	$0.427_{\pm 0.077}$
UWaveGestureLibrary	$0.827_{\pm 0.035}$	$0.823_{\pm 0.033}$	$0.821_{\pm 0.038}$	$0.812_{\pm 0.038}$

Dataset \ Model Component	$\{\boldsymbol{h}_t, \boldsymbol{h}_d\}$	$\{\boldsymbol{h}_t, \boldsymbol{h}_f\}$	$\{\boldsymbol{h}_d, \boldsymbol{h}_f\}$	$\boldsymbol{h}_t$ only	$\boldsymbol{h}_d$ only	$\boldsymbol{h}_f$ only
AtrialFibrillation	$0.293_{\pm 0.038}$	$0.227_{\pm 0.038}$	$0.360_{\pm 0.061}$	$0.240_{\pm 0.077}$	$0.413_{\pm 0.057}$	$0.387_{\pm 0.088}$
BasicMotions	$0.970_{\pm 0.012}$	$0.990_{\pm 0.015}$	$0.995_{\pm 0.012}$	$0.965_{\pm 0.030}$	$0.815_{\pm 0.061}$	$0.990_{\pm 0.015}$
FingerMovements	$0.578_{\pm 0.041}$	$0.552_{\pm 0.023}$	$0.564_{\pm 0.039}$	$0.560_{\pm 0.038}$	$0.566_{\pm 0.010}$	$0.566_{\pm 0.016}$
Heartbeat	$0.722_{\pm 0.013}$	$0.726_{\pm 0.011}$	$0.733_{\pm 0.007}$	$0.738_{\pm 0.016}$	$0.719_{\pm 0.017}$	$0.731_{\pm 0.013}$
MotorImagery	$0.556_{\pm 0.010}$	$0.564_{\pm 0.031}$	$0.560_{\pm 0.026}$	$0.540_{\pm 0.017}$	$0.500_{\pm 0.001}$	$0.580_{\pm 0.036}$
RacketSports	$0.793_{\pm 0.009}$	$0.807_{\pm 0.025}$	$0.680_{\pm 0.046}$	$0.803_{\pm 0.013}$	$0.676_{\pm 0.020}$	$0.593_{\pm 0.037}$
SelfRegulationSCP1	$0.872_{\pm 0.033}$	$0.872_{\pm 0.021}$	$0.813_{\pm 0.017}$	$0.878_{\pm 0.010}$	$0.509_{\pm 0.018}$	$0.810_{\pm 0.013}$
SelfRegulationSCP2	$0.552_{\pm 0.032}$	$0.570_{\pm 0.027}$	$0.598_{\pm 0.017}$	$0.547_{\pm 0.024}$	$0.517_{\pm 0.038}$	$0.601_{\pm 0.008}$
StandWalkJump	$0.413_{\pm 0.088}$	$0.413_{\pm 0.074}$	$0.400_{\pm 0.134}$	$0.373_{\pm 0.061}$	$0.413_{\pm 0.088}$	$0.320_{\pm 0.031}$
UWaveGestureLibrary	$0.813_{\pm 0.036}$	$0.656_{\pm 0.196}$	$0.764_{\pm 0.038}$	$0.529_{\pm 0.346}$	$0.752_{\pm 0.027}$	$0.333_{\pm 0.107}$