META: META-LEARNING ENHANCED TEMPORAL-FREQUENCY ALIGNMENT FOR TIME SERIES DOMAIN ADAPTATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Time series classification (TSC) underpins clinical decision support, yet differences in patient demographics and institutional data collection workflows induce domain shift that severely degrades model performance on the target domain. Unsupervised domain adaptation (UDA) mitigates this by aligning distributions, but we show that the two dominant alignment paradigms each leave a distinct structural blind spot. First, joint maximum mean discrepancy (JMMD) constrains only class-conditional means, leaving mean-preserving drifts in its null space unpenalized, which drifts enlarge intra-class dispersion, directly reducing discriminability. Second, contrastive learning pulls all same-class samples toward a single centroid, and this non-uniform attraction radially stretches each sub-mode, enlarging sub-modal dispersion and destroying the multi-modal structure that carries clinically meaningful heterogeneity. We address both blind spots with META, a unified time-series UDA framework comprising two modules: (i) Time-Frequency Contrastive Learning leverages the temporal dependency of time and the stability of frequency to learn complementary contextual representations. We prove that at its global optimum same-class features collapse to class centers forming a regular simplex. (ii) MetaMMD treats the Gaussian kernel bandwidths of MMD as learnable parameters optimized via bi-level optimization against a domain discrimination loss, which differentiates bandwidths so that same-sub-mode cross-domain kernel weights remain bounded below by a positive constant while cross-sub-mode weights vanish. We prove that this shrinks within-sub-mode variance exponentially and drives sub-modal dispersion toward zero without merging sub-modes. Experiments on eight datasets across four tasks demonstrate that META consistently outperforms state-of-the-art UDA and time-series UDA baselines.

1 INTRODUCTION

Time series classification (TSC) for healthcare has become increasingly important for decision support, as it strengthens diagnostic capabilities Thapa et al. (2026). However, variations in data collection workflows and demographics cause the target distribution to diverge from the source, namely domain shift Choi et al. (2025). Domain shift across devices and populations degrades model generalization, causing algorithmic biases, missed diagnoses, and unnecessary interventions. For example, the AP50 dropped by 70.3% when the imaging devices differed Matsui et al. (2025).

Unsupervised domain adaptation (UDA) Lu et al. (2025) mitigates domain shift by transferring knowledge from labeled to unlabeled domains. A mainstream UDA strategy is constructing probability distribution distance metrics, among which Maximum Mean Discrepancy (MMD) Gretton et al. (2012) is a measure that aligns domains by matching mean embeddings in a Reproducing Kernel Hilbert Space (RKHS). Beyond marginal moment matching Chen et al. (2020), Joint Maximum Mean Discrepancy (JMMD) Wang et al. (2026) aligns conditional distributions via Class-wise Maximum Mean Discrepancy (CMMD) Ge et al. (2020), while Weighted Class-wise Maximum Mean Discrepancy (WCMMD) Huang et al. (2026) introduces class weights to handle imbalance. With its solid theoretical foundation, JMMD has been applied in diverse domains, including medical image segmentation for aligning clustered feature representations Li et al. (2025).

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

Despite its rapid progress, JMMD constrains only the class-conditional mean, leaving mean-preserving feature drifts in its null space unconstrained. A sufficiently large drift can therefore enlarge intra-class feature dispersion, the spread of same-class features in the embedding space. Specifically, such a drift leaves both the class mean and the JMMD loss unchanged, while strictly increasing the Frobenius norm of the class covariance. Contrastive learning explicitly compresses intra-class dispersion: by pulling same-class positive pairs toward a common class center on the unit hypersphere, it collapses the intra-class covariance to zero at its optimum Wu et al. (2026); Lin et al. (2025b); Koromilas et al. (2024). Recent efforts have adapted contrastive learning to time-series UDA Jia et al. (2025); Darban et al. (2025), yet they mainly design time-domain augmentations, overlooking the more stable frequency-domain information that is less sensitive to domain shifts.

However, the global attraction in contrastive learning can enlarge sub-modal dispersion, the internal spread of each sub-mode within a class, which arises when same class target samples form multiple clusters due to device variations. Because this attraction pulls all same class samples toward a single centroid, samples farther from the centroid are pulled more strongly, inducing a proportional stretching along each sub-mode's centered directions that increases the Frobenius norm of the sub-modal covariance. Such stretching inflates sub-modal dispersion and destroys inter-sub-mode separation, degrading classification performance. Therefore, an effective time-series UDA method need to 1) align conditional distributions with low intra-class dispersion while 2) preserving sub-modal structures with minimal sub-modal dispersion.

To address these challenges, we propose META (Meta-learning Enhanced Temporal-Frequency Alignment), a unified framework that minimizes intra-class dispersion and sub-modal dispersion. Figure 1 illustrates the geometric intuition behind these two designs on the unit hypersphere. First, since JMMD constrains only the class-conditional mean, we introduce Time-Frequency Contrastive Learning, which constructs positive pairs from time- and frequency-augmented views, enabling mutual learning between temporal and spectral features that pulls same-class features together. Temporal features capture sequential dependencies while spectral magnitudes are robust to domain shift, so their mutual learning guides the encoder toward semantically shared representations. Second, since global contrastive learning pulls same-class samples toward a single centroid and thereby stretches each sub-mode, we propose MetaMMD, which adaptively optimizes the Gaussian kernel bandwidths of MMD via bi-level optimization: the inner loop updates the base model with the current bandwidths, while the outer loop updates the bandwidth generator using the domain discrimination loss. This differentiates the bandwidths so that within-sub-mode kernel weights remain bounded below by a positive constant while cross-sub-mode weights vanish, shrinking the within-sub-mode variance exponentially. The contributions of META are as follows:

- We propose Time-Frequency Contrastive Learning to reduce intra-class dispersion through mutual learning between time and frequency augmented views. Supervised contrastive learning on the source domain pulls same-class features together, and unsupervised contrastive learning on the target maintains intra-class compactness. Theorem 4.1 proves that at the global optimum same-class features collapse to a single class center while different class centers form a regular simplex on the hypersphere.

- We propose Cross-Domain Contrastive Learning to align source and target contextual representations. Exploiting the cross-domain stability of frequency features, it retrieves the most similar source sample for each target sample, treating the matched pair as a positive and the remaining source samples as negatives.

- We propose MetaMMD to counter the sub-modal stretching induced by global contrastive learning. By adaptively optimizing the Gaussian kernel bandwidths via bi-level optimization, it assigns vanishing kernel weights to cross-sub-mode pairs while preserving positive weights for within-sub-mode pairs, so that each sub-mode is aligned locally without destroying the multi-modal structure. We prove that the bi-level optimization induces this bandwidth differentiation and that the within-sub-mode variance then shrinks exponentially (Lemma 4.1), minimizing sub-modal dispersion. Experiments on four tasks across eight datasets demonstrate the effectiveness of META.

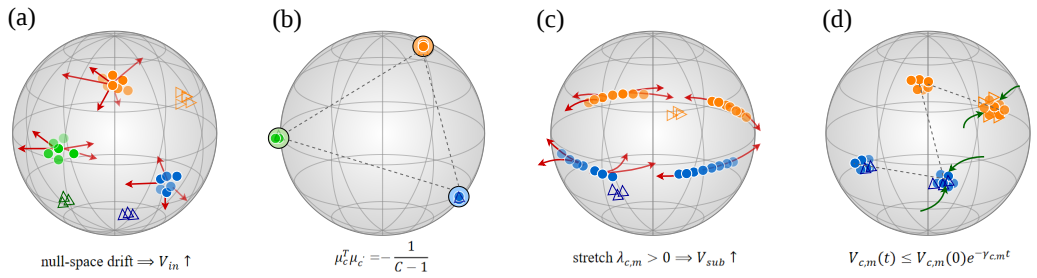


Figure 1: META motivation on the unit hypersphere: (a) JMMD enlarges V_{in}; (b) contrastive learning collapses classes; (c) but stretches sub-modes; (d) MetaMMD preserves multi-modal structures.

2 Preliminaries

2.1 Problem Definition

Definition 2.1 (TSC-UDA). *Given a labeled source domain $\mathcal{D}_s = \{(\mathbf{x}_i^s, y_i^s)\}_{i=1}^{n_s}$ and an unlabeled target domain $\mathcal{D}_t = \{\mathbf{x}_j^t\}_{j=1}^{n_t}$, where each $\mathbf{x} \in \mathbb{R}^{T\times d}$ is time series with T time steps and d channels, and $y_i \in \mathcal{C}$, TSC-UDA aims to learn a mapping $f : \mathbb{R}^{T\times d} \to \mathcal{C}$ that predicts target labels.*

Definition 2.2 (Intra-Class Dispersion V_{in}). *To characterize the class-conditional compactness of the feature space, we define Intra-class Dispersion based on GMM. Let $\mathbf{z} = \phi(\mathbf{x}) \in \mathbb{R}^D$ be the embedding. For each class c, let $\mathcal{S}_c = \{\mathbf{z}_i^s \mid y_i^s = c\}$, $\boldsymbol{\mu}_c = \frac{1}{|\mathcal{S}_c|}\sum_{\mathbf{z}\in\mathcal{S}_c}\mathbf{z}$, $\boldsymbol{\Sigma}_c = \frac{1}{|\mathcal{S}_c|}\sum_{\mathbf{z}\in\mathcal{S}_c}(\mathbf{z}-\boldsymbol{\mu}_c)(\mathbf{z}-\boldsymbol{\mu}_c)^\top$. Then $V_{in} = \sum_{c=1}^{C}\alpha_c\|\boldsymbol{\Sigma}_c\|_F$, with $\alpha_c = |\mathcal{S}_c|/\sum_{c'}|\mathcal{S}_{c'}|$. A larger V_{in} indicates more dispersed and overlapping classes with reduced discriminability.*

Definition 2.3 (Sub-Modal Dispersion V_{sub}). *To measure the internal compactness of each sub-mode in time series, we define Sub-Modal Dispersion, where a class c may contain M_c sub-modes due to individual differences. Let $\mathcal{S}_{c,m}$ be the embedded samples of the m-th sub-mode of class c, and let $\boldsymbol{\mu}_{c,m}$ and $\boldsymbol{\Sigma}_{c,m}$ be its mean and covariance defined analogously to 2.2. Then $V_{sub} = \sum_{c=1}^{C}\sum_{m=1}^{M_c}\beta_{c,m}\|\boldsymbol{\Sigma}_{c,m}\|_F$, where $\beta_{c,m} = |\mathcal{S}_{c,m}|/\sum_{c',m'}|\mathcal{S}_{c',m'}|$. Unlike V_{in}, V_{sub} measures within-sub-mode compactness, reflecting whether the intrinsic multi-modal structure is preserved.*

2.2 Limitation of JMMD on Intra-Class Feature Dispersion

JMMD aligns joint distributions via mean embeddings in RKHS, yet its class-wise form only constrains first-order moments, so V_{in} can increase without changing the JMMD loss as:

$$\mathcal{L}_{JMMD} = \left\| \frac{1}{n_s}\sum_{i=1}^{n_s}\psi(\mathbf{x}_i^s)\otimes\phi(y_i^s) - \frac{1}{n_t}\sum_{j=1}^{n_t}\psi(\mathbf{x}_j^t)\otimes\phi(\hat{y}_j^t)\right\|_{\mathcal{H}}^2, \tag{1}$$

where ψ and ϕ are the feature and label mappings to RKHS, and $\hat{y}_j^t$ is pseudo-label of $\mathbf{x}_j^t$.

The class-wise JMMD depends only on the class-conditional mean. Thus any mean-preserving perturbation $\Delta_{c,0}$ with $\Delta_{c,0}^\top \mathbf{1}_{n_c} = 0$ lies in its null space, leaving both the mean and $\mathcal{L}_{JMMD}$ unchanged and unpenalized. If other optimization forces scale such a drift beyond an explicit threshold T_c, the class-conditional covariance provably enlarges, and the intra-class dispersion increases. Detailed proof and assumption are provided in Appendix A.

Lemma 2.1 (JMMD Null-Space Drift Increases V_{in}). *Let $\mathbf{Z}_c$ be the centered feature matrix, $\mathbf{Z}_c^\top \mathbf{1}_{n_c} = 0$, and $\boldsymbol{\Sigma}_c = n_c^{-1}\mathbf{Z}_c^\top\mathbf{Z}_c$. Suppose optimization produces a nonzero mean-preserving drift $\Delta_{c,0} \neq 0$ with $\Delta_{c,0}^\top \mathbf{1}_{n_c} = 0$, and set $\mathbf{Z}_c(t_c) = \mathbf{Z}_c + t_c\Delta_{c,0}$. Define $q_c(t) := \|\boldsymbol{\Sigma}_c(t)\|_F^2 - \|\boldsymbol{\Sigma}_c\|_F^2$ and $T_c := \max\{|r| : q_c(r) = 0\}$. Here T_c is the threshold such that for all $|t_c| > T_c$, $\|\boldsymbol{\Sigma}_c(t_c)\|_F > \|\boldsymbol{\Sigma}_c\|_F$. If $|t_c| > T_c$ holds for every class c, then*

$$V_{in}' = \sum_{c=1}^{C}\alpha_c\|\boldsymbol{\Sigma}_c(t_c)\|_F > \sum_{c=1}^{C}\alpha_c\|\boldsymbol{\Sigma}_c\|_F = V_{in}. \tag{2}$$

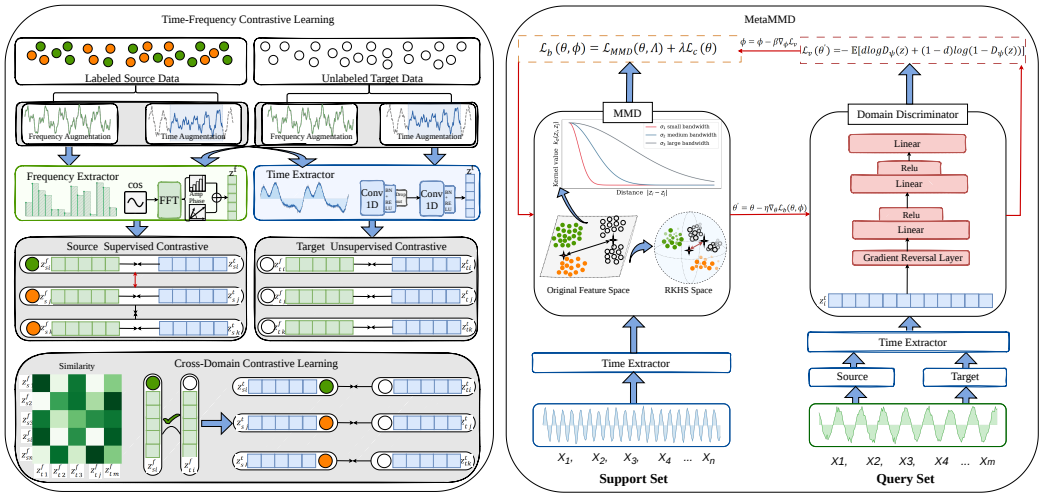


Figure 2: META: (a) Time-Frequency Contrastive Learning augments time/frequency views and applies supervised, unsupervised, and cross-domain contrastive learning. (b) MetaMMD generates adaptive bandwidths via bi-level optimization. Blue/Red arrows denote data/parameter flow.

3 METHODOLOGY

3.1 OVERVIEW

In this section, we introduce META, a unified framework designed to address the challenges of intra-class dispersion and sub-modal destruction in time series. The overall framework of META is illustrated in Figure 2, which consists of Time-Frequency Contrastive Learning and MetaMMD:

- To reduce intra-class dispersion, META introduces Time-Frequency Contrastive Learning, which leverages complementary time-frequency views with supervised source contrast and unsupervised target contrast to enhance intra-class compactness. For domain alignment, cross-domain contrastive learning retrieves similar source samples via more stable frequency features (See Section 3.2).
- To minimize sub-modal dispersion, META further introduces MetaMMD, which adaptively aligns sub-modal structures through bi-level optimization that dynamically adjusts the Gaussian kernel bandwidths. The inner loop updates model with current bandwidths, while the outer loop refines the bandwidths via domain discrimination. Specifically, the bandwidth optimization prevents cross sub-modal collapse (See Section 3.4).

3.2 TIME-FREQUENCY CONTRASTIVE LEARNING

Time-Frequency Contrastive Learning aims to pull semantically similar samples closer to compress V_{in}. Since time series exhibit sequential dependency but domain shift involves both time-domain perturbations and spectral variations, time-domain augmentations alone are vulnerable to spectral changes. META therefore introduces frequency-domain augmentations and enforces time-frequency consistency to learn domain-invariant features. Specifically, time-domain augmentation randomly applies one of three strategies to yield $\tilde{x}^t$: random cropping, random time masking, or additive Gaussian noise. Frequency-domain augmentation masks the FFT magnitude spectrum and applies inverse FFT to yield $\tilde{x}^f$. Then $\tilde{x}^t$ is fed into the temporal extractor E_t as z^t, which is used for classification, while $\tilde{x}^f$ is fed into frequency extractor E_f to extract $z^f = [a_i, p_i]$, where a_i and p_i denote magnitude and phase. Projection heads map z^t and z^f, yielding L2-normalized z_p^t and z_p^f.

Supervised Contrast on Source Domain To reduce V_{in}, we employ supervised contrastive learning on $\mathcal{D}_s$, pulling same-class features together while pushing different-class features apart. For B_s source samples, each has two views $\tilde{z}_i^t$ and $\tilde{z}_i^f$, yielding $2B_s$ views in $\mathcal{I}$. The positive set $\mathcal{P}(i)$

includes all views with the same label as anchor i, including its other view. The loss is:

$$\mathcal{L}_{\text{SupCon}} = -\sum_{i\in\mathcal{I}} \frac{1}{|\mathcal{P}(i)|} \sum_{p\in\mathcal{P}(i)} \log \frac{\exp(\tilde{z}_i \cdot \tilde{z}_p/\tau)}{\sum_{j\in\mathcal{I}\setminus\{i\}} \exp(\tilde{z}_i \cdot \tilde{z}_j/\tau)}, \tag{3}$$

where τ scales the similarity distribution, compressing V_{in} toward its theoretical minimum.

Unsupervised Contrast on Target Domain For the unlabeled target domain $\mathcal{D}_t$, we pull the time-frequency views of the same sample together while pushing apart views of different samples. Let $\tilde{z}_k^t$ and $\tilde{z}_k^f$ be the projected time and frequency features of target sample k, forming a positive pair. For B_t target samples, this yields $2B_t$ views. For an anchor $\tilde{z}_k$, $\tilde{z}_{\text{pos}(k)}$ is its other view, and all remaining views are negatives. The loss $\mathcal{L}_{\text{TCon}}$ is:

$$\mathcal{L}_{\text{TCon}} = -\frac{1}{2B_t} \sum_{k=1}^{2B_t} \log \frac{\exp(\tilde{z}_k \cdot \tilde{z}_{\text{pos}(k)}/\tau)}{\sum_{a\neq k} \exp(\tilde{z}_k \cdot \tilde{z}_a/\tau)}. \tag{4}$$

Cross-Domain Contrast To align the domains, we use the cross-domain stability of frequency to retrieve the most similar source sample for each target sample and pull their time features closer. Given frequency feature matrices Z_s^f and Z_t^f, we compute similarity $\text{sim}(i,j) = Z_t^f[i] \cdot (Z_s^f[j])^\top$ and retrieve $\text{idx}_i = \arg\max_j \text{sim}(i,j)$. The positive pair consists of the target time feature $\tilde{z}_i^t$ and the retrieved source time feature $\tilde{z}_{\text{idx}_i}^t$, where $\tilde{z}_i^t$ and $\tilde{z}_j^t$ denote the projected time features of target sample i and source sample j, respectively. The cross-domain contrastive loss as:

$$\mathcal{L}_{\text{Cross}} = -\frac{1}{B_t} \sum_{i=1}^{B_t} \log \frac{\exp(\tilde{z}_i^t \cdot \tilde{z}_{\text{idx}_i}^t/\tau)}{\sum_{j=1}^{B_s} \exp(\tilde{z}_i^t \cdot \tilde{z}_j^t/\tau)}. \tag{5}$$

Finally, the overall contrastive loss is defined as $\mathcal{L}_{\text{Con}} = \lambda_1\mathcal{L}_{\text{SupCon}} + \lambda_2\mathcal{L}_{\text{TCon}} + \lambda_3\mathcal{L}_{\text{Cross}}$.

3.3 Limitation of Global Contrastive Learning on Multi-Modal Distribution

Global contrastive learning pulls same-class target samples toward a common centroid, but non-uniformly stretches each sub-mode along its centered feature direction, enlarging the sub-modal covariance Frobenius norm and thus increasing V_{sub}. Intuitively, samples lying farther from the sub-mode centroid receive a proportionally larger displacement toward the class centroid, so the sub-mode is radially stretched while being translated. Formally, for sub-mode m of class c, write the displacement of sample i as $\delta_i = \Delta\boldsymbol{\mu}_{c,m} + \lambda_{c,m}u_i + \epsilon_i$, where $u_i = z_i^o - \boldsymbol{\mu}_{c,m}^o$ is the original centered feature, $\Delta\boldsymbol{\mu}_{c,m}$ is the sub-modal mean drift, $\lambda_{c,m} > 0$ is the stretching coefficient, and ϵ_i is a zero-mean residual with $\sum_i \epsilon_i = 0$ and $\sum_i u_i\epsilon_i^\top = \mathbf{0}$. The new centered feature is then $u_i' = (1+\lambda_{c,m})u_i + \epsilon_i$, whose covariance satisfies $\boldsymbol{\Sigma}_{c,m}' = (1+\lambda_{c,m})^2\boldsymbol{\Sigma}_{c,m} + \frac{1}{n_{c,m}}\sum_i \epsilon_i\epsilon_i^\top \succeq (1+\lambda_{c,m})^2\boldsymbol{\Sigma}_{c,m}$, which strictly enlarges the sub-modal covariance (proof in Appendix C):

Proposition 3.1. *The displaced sub-modal covariance satisfies* $\|\boldsymbol{\Sigma}_{c,m}'\|_F \geq (1+\lambda_{c,m})^2\|\boldsymbol{\Sigma}_{c,m}\|_F > \|\boldsymbol{\Sigma}_{c,m}\|_F$ *whenever* $\boldsymbol{\Sigma}_{c,m} \neq \mathbf{0}$. *Then, with* $\beta_{c,m} > 0$,

$$V_{sub}' = \sum_{c=1}^{C}\sum_{m=1}^{M_c} \beta_{c,m}\|\boldsymbol{\Sigma}_{c,m}'\|_F > \sum_{c=1}^{C}\sum_{m=1}^{M_c} \beta_{c,m}\|\boldsymbol{\Sigma}_{c,m}\|_F = V_{sub}. \tag{6}$$

3.4 MetaMMD

MetaMMD learns Gaussian bandwidths via bi-level optimization to shift MMD from global to sub-modal alignment, preserving sub-modal independence. Let f_θ be the base model, g_ϕ be the meta-model that outputs a bandwidth σ_{ij} for each pair of embeddings to form the bandwidth matrix Λ, and D_ψ be the domain discriminator for the outer-loop domain loss. Source and target samples form the support set $\mathcal{S}_\tau$ and query set $\mathcal{Q}_\tau$. θ, ϕ, ψ are randomly or pre-trained initialized.

Inner Loop: Base Model Update on Support Set The inner loop updates f_θ on $\mathcal{S}_\tau$ by minimizing a combined loss for domain alignment and class discriminability. Samples in $\mathcal{S}_\tau$ are passed

through f_θ to obtain embeddings $\{z_i\}$. The meta-model g_ϕ outputs a bandwidth σ_{ij} for each pair, forming the bandwidth matrix Λ. The inner loss combines MMD and source classification $\mathcal{L}_b(\theta, \phi) = \mathcal{L}_{\text{MMD}}(\theta, \Lambda) + \lambda_{cls}\mathcal{L}_{cls}(\theta) + \mathcal{L}_{\text{Con}}$, where the adaptive MMD loss is computed as:

$$\mathcal{L}_{\text{MMD}}(\theta, \Lambda) = \frac{1}{n_s^2}\sum_{i,j} k_{\sigma_{ij}}(z_{s,i}, z_{s,j}) + \frac{1}{n_t^2}\sum_{i,j} k_{\sigma_{ij}}(z_{t,i}, z_{t,j}) - \frac{2}{n_s n_t}\sum_{i,j} k_{\sigma_{ij}}(z_{s,i}, z_{t,j}), \quad (7)$$

with the kernel function defined as $k_{\sigma_{ij}}(\mathbf{z}_i, \mathbf{z}_j) = \exp\left(-\|\mathbf{z}_i - \mathbf{z}_j\|^2/2\sigma_{ij}^2\right)$. Finally, the base model parameters are updated via $\theta' = \theta - \eta\nabla_\theta\mathcal{L}_b(\theta, \phi)$, where η is the inner-loop learning rate.

Outer Loop: Bandwidth Update on Query Set The outer loop updates the meta-model parameters ϕ on $\mathcal{Q}_\tau$ by maximizing $\mathcal{L}_v$, thereby refining the bandwidth generation strategy to enable sub-modal alignment. With the updated $f_{\theta'}$, we extract features from $\mathcal{Q}_\tau$ and compute the outer loss:

$$\mathcal{L}_v(\theta', \psi) = -\mathbb{E}_{(z,d)\sim\mathcal{Q}_\tau}\left[d\log D_\psi(z) + (1-d)\log(1 - D_\psi(z))\right], \quad (8)$$

where $d \in \{0, 1\}$ denotes the domain label. Since ϕ affects $\mathcal{L}_v$ only indirectly through θ', computing $\nabla_\phi\mathcal{L}_v$ directly would require unrolling the inner-loop gradient chain, which is computationally prohibitive. We therefore adopt a DARTS-style finite-difference approximation:

$$\nabla_\phi\mathcal{L}_v \approx \frac{\eta}{2\epsilon}\left[\nabla_\phi\mathcal{L}_b(\theta^-, \phi) - \nabla_\phi\mathcal{L}_b(\theta^+, \phi)\right], \quad (9)$$

where $g_v = \nabla_{\theta'}\mathcal{L}_v(\theta')$ is computed on $\mathcal{Q}_\tau$, $\theta^\pm = \theta' \pm \epsilon g_v$, and the two gradients of the inner loss $\mathcal{L}_b$ are evaluated on $\mathcal{S}_\tau$ with θ^+ and θ^-. Finally, the meta-model ascends the outer loss to maximize domain confusion, $\phi \leftarrow \phi + \beta\nabla_\phi\mathcal{L}_v$, while the discriminator descends it to sharpen domain discrimination, $\psi \leftarrow \psi - \gamma\nabla_\psi\mathcal{L}_v$, where β and γ are the outer-loop and discriminator learning rates, respectively. Training alternates between inner and outer updates until convergence.

4 Theoretical Analysis

4.1 Time-Frequency Contrastive Learning Compresses Intra-Class Dispersion

We prove that the supervised contrastive learning within Time-Frequency Contrastive Learning compresses intra-class dispersion at its global optimum. At this optimum, same-class features collapse to a single center on the unit hypersphere while different class centers are uniformly separated, driving the intra-class covariance to zero and minimizing V_{in} (proved in Appendix B).

Theorem 4.1. *Suppose $\mathcal{L}_{SupCon}$ attains its global optimum. Then: (i) for each class $c \in \mathcal{C}$, all normalized features of class c collapse to the same class center: $\tilde{\mathbf{z}}_i = \boldsymbol{\mu}_c$ for all $i \in \mathcal{S}_c$, with $\|\boldsymbol{\mu}_c\|_2 = 1$; (ii) the intra-class covariance matrix vanishes, $\boldsymbol{\Sigma}_c = \frac{1}{|\mathcal{S}_c|}\sum_{i\in\mathcal{S}_c}(\tilde{\mathbf{z}}_i - \boldsymbol{\mu}_c)(\tilde{\mathbf{z}}_i - \boldsymbol{\mu}_c)^\top = \mathbf{0}$, and consequently*

$$\|\boldsymbol{\Sigma}_c\|_F = 0, \qquad V_{in} = \sum_{c=1}^{C}\alpha_c\|\boldsymbol{\Sigma}_c\|_F = 0; \quad (10)$$

(iii) if the batch is class-balanced ($n_c = m$ for all c) and $C \le D + 1$, the class centers minimizing the uniformity term form a regular simplex on $\mathbb{S}^{D-1}$: $\boldsymbol{\mu}_c^\top\boldsymbol{\mu}_{c'} = -\frac{1}{C-1}$ and $\sum_{c=1}^{C}\boldsymbol{\mu}_c = \mathbf{0}$.

Discussion. Theorem 4.1 shows that at the global optimum of $\mathcal{L}_{\text{SupCon}}$, same-class features collapse to a single center, and the class centers form a regular simplex with pairwise inner product $-1/(C-1)$, the maximally separated configuration of C unit vectors. This simplex is strictly better than orthogonal centers, since $e^{-1/(\tau(C-1))} < e^0$. Orthogonality arises only when negative inner products are forbidden, or approximately when $C = D + 1$ and D is large so that $-1/(C-1) \approx 0$. Smaller τ widens this gap and strengthens the collapse.

4.2 MetaMMD Bandwidth Differentiation Shrinks Sub-Modal Dispersion

The bi-level optimization induces bandwidth differentiation that compacts each sub-mode. Since the outer loop maximizes domain confusion, the kernel-gradient weight evolves as $\dot{w}_{ij} = \beta\eta c_{ij}q_{ij}[w'(\sigma_{ij})]^2$, whose sign aligns the pair's inner-loop displacement with the adversarial direction. Same-sub-mode cross-domain weights reach their maxima, while within-domain and cross-sub-mode weights vanish, yielding $w^s_{\min} \le w^s_{ij} \le w^s_{\max}$ and $w_{ij} \le \bar{w}$, with $w^s_{\min} > 2\bar{w}$.

Lemma 4.1. *For sub-mode m of class c, let the target samples be $z_1, \dots, z_{n_{c,m}}$ with covariance $\mathbf{\Sigma}_{c,m}$ and $V_{c,m} = \mathrm{tr}(\mathbf{\Sigma}_{c,m})$ (an upper bound on $\|\mathbf{\Sigma}_{c,m}\|_F$ since $\mathbf{\Sigma}_{c,m} \succeq \mathbf{0}$), and let the matched source sub-mode have mean $\boldsymbol{\mu}^s_{c,m}$ and satisfy $\frac{1}{n^s_{c,m}} \sum_j \|z_{s,j} - \boldsymbol{\mu}^s_{c,m}\|^2 \le \rho^2$. Under the weight configuration induced by Lemma D.1, the inner flow satisfies, with $\gamma_{c,m} = \frac{4\eta}{n_{c,m}}(w^s_{\min} - 2\bar{w}) > 0$ and $M_{c,m}(t) = \frac{1}{n_{c,m}} \sum_i \|z_i(t) - \boldsymbol{\mu}^s_{c,m}\|^2$,*

$$V_{c,m}(t) \;\le\; \left(e^{-\gamma_{c,m} t/2} \sqrt{M_{c,m}(0)} + \frac{w^s_{\max}\, \rho}{w^s_{\min} - 2\bar{w}} \right)^2 . \tag{11}$$

In particular, if the source sub-mode has collapsed ($\rho = 0$), then $V_{c,m}(t) \le M_{c,m}(0) e^{-\gamma_{c,m} t} \to 0$, hence $\|\mathbf{\Sigma}_{c,m}(t)\|_F \to 0$ for every sub-mode and $V_{sub}(t) \to 0$ (proved in Appendix D).

Discussion. Lemma 4.1 shows that MetaMMD compacts each sub-mode: bandwidth differentiation turns the MMD gradient field into a net contraction, where matched cross-domain pairs attract with weight at least $w^s_{\min}$ while within-domain repulsion is damped below $\bar{w}$, and $w^s_{\min} > 2\bar{w}$ ensures attraction dominates, so the sub-modal variance decays exponentially at rate $\gamma_{c,m}/2$ toward a floor set by source compactness. When $\rho = 0$, this drives $V_{sub} \to 0$, and vanishing cross-sub-mode weights preserve sub-mode independence.

5 Experiments

We conduct experiments to compare META against various baseline methods over eight datasets across four tasks to answer the following **Research Questions (RQs)**: 1) **RQ1:** Is META an effective time-series UDA method? 2) **RQ2:** Does the Time-Frequency Contrastive Learning effectively compress intra-class dispersion V_{in}? 3) **RQ3:** Does the MetaMMD minimize sub-modal dispersion V_{sub}? Implementation details are deferred to Appendix F.

5.1 Experiment Setup

Datasets: We evaluate META on 1) human activity recognition (HAR) with HHAR_SA, HHAR_D Stisen et al. (2015), WISDM Kwapisz et al. (2011), HAR Garcia-Gonzalez et al. (2020), 2) sleep stage classification (SSC) with Sleep-EDF, CAP Goldberger et al. (2000), 3) gesture recognition (GR) with EMG Lu et al. (2022), and 4) motor imagery (MI) with PCL Cho et al. (2017).

Baselines: We compare META against 5 UDA methods including CDAN Long et al. (2018), HoMM Chen et al. (2020), MMDA Rahman et al. (2020), DSAN Zhu et al. (2020), JMMD-HSIC Wang et al. (2026) and 10 time-series UDA methods including CoDATS Wilson et al. (2020), RAINCOAT He et al. (2023), AdvSKM Liu & Xue (2021), CALDA Wilson et al. (2023), CLUDA Ozyurt et al. (2022), CoTMix Jia et al. (2025), ConFGD Guo et al. (2025), ACON Liu et al. (2024), GLA-DA Tu et al. (2024) and EDEN Liu et al. (2026), with Source results reported for domain generalization difficulty. For each dataset, we report average accuracy and Macro-F1 over random domain pairs, with standard deviation across pairs indicating robustness.

5.2 Effectiveness of META (RQ-1)

Results Table 1 reports the mean accuracy across all datasets (full results in Appendix G). META achieves average improvements of 3.3% on HAR , 7.9% on SSC, 1.8% on GR, and 1.7% MI tasks.

Ablation Study To assess each component, we ablate Time-Frequency Contrastive Learning (*w/o C.*), MetaMMD (*w/o M.*), both (*w/o C.M.*), and each contrastive loss. As shown in Table 2, META achieves the best performance on all datasets. Among the contrastive losses (CL), removing $\mathcal{L}_{\text{SupCon}}$, $\mathcal{L}_{\text{TCon}}$, and $\mathcal{L}_{\text{Cross}}$ causes drops on accuracy, with $\mathcal{L}_{\text{SupCon}}$ being the most critical.

5.3 Analysis of Time-Frequency Contrastive Learning (RQ-2)

Analysis of V_{in} To analyze intra-class dispersion V_{in}, Figure 4a compares HoMM (JMMD), META without MetaMMD (*w/o M*), and full META. The results show that *w/o M* significantly reduces V_{in}

Table 1: Mean and standard deviation of accuracies across eight datasets over domain-pairs and 5 independent runs. The **best** results are in bold, and the second-best results are underlined.

Task	HAR				SSC		GR	MI
Method	HHAR_SA	HHAR_D	WISDM	HAR	Sleep-EDF	CAP	EMG	PCL
Source	56.9 ± 19.1	47.3 ± 15.6	62.4 ± 12.5	81.6 ± 8.5	57.0 ± 9.2	61.7 ± 9.5	73.7 ± 6.4	60.3 ± 3.7
CDAN	73.4 ± 22.9	57.8 ± 14.2	69.9 ± 13.1	94.4 ± 7.0	75.4 ± 7.8	68.0 ± 7.1	78.0 ± 4.8	61.7 ± 3.5
HoMM	66.1 ± 22.0	55.4 ± 14.3	68.1 ± 9.6	90.7 ± 8.1	74.2 ± 9.1	65.9 ± 8.3	77.9 ± 4.0	63.2 ± 4.5
MMDA	67.8 ± 26.0	53.2 ± 14.6	75.0 ± 7.4	89.2 ± 13.8	74.7 ± 8.6	63.6 ± 10.0	78.2 ± 4.2	52.3 ± 3.2
DSAN	71.4 ± 25.6	53.1 ± 14.6	73.0 ± 9.1	81.6 ± 15.5	72.3 ± 8.3	66.3 ± 8.1	78.2 ± 4.8	62.1 ± 4.3
JMMD-HSIC	64.3 ± 23.1	51.8 ± 14.1	70.4 ± 11.6	89.6 ± 10.0	74.7 ± 7.9	64.7 ± 8.7	77.0 ± 4.0	63.6 ± 4.1
CoDATS	73.6 ± 22.9	55.6 ± 15.5	73.6 ± 12.5	92.3 ± 10.4	71.1 ± 7.2	69.3 ± 8.3	77.8 ± 5.7	63.5 ± 4.5
RAINCOAT	74.5 ± 21.8	53.8 ± 16.8	76.1 ± 12.1	94.1 ± 7.3	75.2 ± 7.8	63.3 ± 9.2	78.0 ± 5.2	61.4 ± 4.4
AdvSKM	61.1 ± 21.4	50.8 ± 15.1	70.0 ± 8.7	88.4 ± 10.3	69.3 ± 10.7	64.5 ± 8.4	74.7 ± 5.6	63.8 ± 4.2
CALDA	70.6 ± 19.1	55.7 ± 15.7	74.2 ± 18.9	92.0 ± 9.6	74.7 ± 10.1	69.5 ± 6.0	65.9 ± 4.7	57.1 ± 5.4
CLUDA	69.1 ± 18.4	53.8 ± 15.4	73.7 ± 11.9	93.3 ± 9.7	73.1 ± 10.3	64.3 ± 6.6	77.5 ± 4.3	61.9 ± 7.5
CoTMix	66.8 ± 12.5	49.2 ± 18.4	73.4 ± 15.7	82.2 ± 11.9	70.4 ± 8.8	58.5 ± 10.9	75.0 ± 4.3	53.1 ± 5.7
ConFGD	80.3 ± 19.0	48.5 ± 15.3	80.7 ± 11.6	96.1 ± 4.6	52.6 ± 10.6	61.9 ± 12.5	77.3 ± 4.4	63.1 ± 4.5
ACON	79.2 ± 19.4	63.7 ± 16.3	84.1 ± 11.1	95.5 ± 6.4	66.9 ± 10.9	72.1 ± 5.5	82.9 ± 5.9	63.1 ± 3.9
GLA-DA	68.1 ± 21.6	47.6 ± 16.6	66.7 ± 7.0	89.6 ± 9.9	72.9 ± 8.3	63.1 ± 10.3	76.1 ± 4.3	63.1 ± 4.3
EDEN	82.7 ± 16.2	68.8 ± 9.8	85.3 ± 13.2	97.4 ± 5.0	68.6 ± 9.3	59.1 ± 13.8	82.2 ± 4.9	57.0 ± 4.6
META	**83.2 ± 16.8**	**72.1 ± 15.7**	**90.7 ± 10.6**	**98.8 ± 2.8**	**83.6 ± 4.1**	**75.6 ± 5.7**	**84.4 ± 4.5**	**64.9 ± 4.1**
Improve(%)	**0.6%**	**4.8%**	**6.3%**	**1.4%**	**10.9%**	**4.9%**	**1.8%**	**1.7%**

Table 2: Ablation study on module and contrastive loss components.

Dataset	Module Ablation				CL Ablation			
	META	*w/o C.*	*w/o M.*	*w/o C.M.*	**CL**	*w/o* $\mathcal{L}_{SupCon}$	*w/o* $\mathcal{L}_{TCon}$	*w/o* $\mathcal{L}_{Cross}$
HHAR_SA	**83.2**	67.9	74.7	56.9	**74.7**	69.2	73.3	71.0
HHAR_D	**72.1**	54.0	62.6	47.3	**62.6**	57.0	60.8	57.4
WISDM	**90.7**	84.3	80.3	62.4	**80.3**	74.2	74.6	75.1
HAR	**98.8**	88.1	93.8	81.6	**93.8**	90.9	88.8	92.2
Sleep-EDF	**83.6**	78.3	77.7	57.0	**77.7**	73.1	77.6	77.3
EMG	**84.4**	76.6	74.6	73.7	**74.6**	61.3	67.5	72.3
Δ Avg.	–	↓10.6	↓8.2	↓22.3	–	↓6.3	↓3.5	↓3.1

across the four datasets, while full META also maintains low V_{in}. Figure 4c further illustrates the evolution on WISDM $7 \rightarrow 30$: as the epoch increases , JMMD's V_{in} steadily rises, whereas META's V_{in} rapidly decreases and remains at a low level.

Time vs. Frequency Augmentations To verify the effectiveness of time-domain and frequency-domain augmentations and their mutual learning, we compare three variants on WISDM: Time-Time, Frequency-Frequency, and Time-Frequency. Time-Frequency achieves the best accuracy as 90.71% outperforming Time-Time 87.33% and Frequency-Frequency 87.68%.

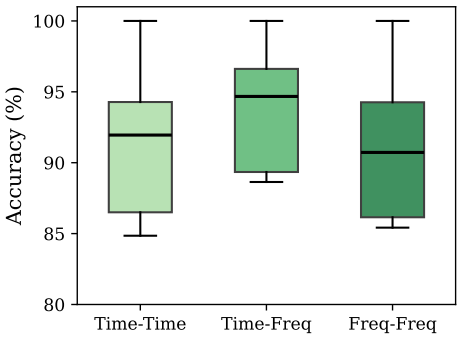
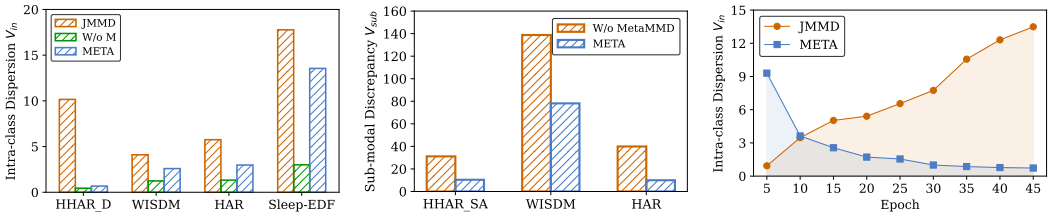


Figure 3: Comparison of Time-Time, Time-Frequency, and Frequency-Frequency enhancement methods on WISDM.

5.4 Analysis of MetaMMD (RQ-3)

Analysis of V_{sub} To evaluate sub-modal dispersion V_{sub}, Figure 4b compares META with the variant without MetaMMD (*w/o M*). META achieves lower V_{sub} than *w/o M*, which demonstrates that it effectively reduces sub-modal dispersion.

Adaptive vs. Heuristic Bandwidth To evaluate the effectiveness of adaptive bandwidth, we compare it with a heuristic bandwidth baseline. The heuristic baseline sets the RBF bandwidth using the mean pairwise squared distance over all source and target samples. As shown in Table 3, adaptive bandwidth consistently achieves higher accuracy across all datasets, demonstrating the benefit of meta-learned bandwidth selection.



(a) Intra-class dispersion V_{in} (b) Sub-modal discrepancy V_{sub} (c) Evolution of V_{in} on WISDM

Figure 4: Analyses of intra-class and sub-modal dispersion. (a) Intra-class feature dispersion V_{in} under JMMD, *w/o M*, and META. (b) Sub-modal dispersion V_{sub} under *w/o M* and META. (c) Evolution of intra-class dispersion V_{in} over training epochs for JMMD and META.

6 Related Work

6.1 Unsupervised Domain Adaptation for Time Series

Table 3: Adaptive and heuristic bandwidth.

Dataset	Adaptive	Heuristic
HHAR_SA	83.2	78.9
WISDM	90.7	83.9
HAR	98.8	94.6
Sleep-EDF	83.6	76.3

Unsupervised Domain Adaptation (UDA) transfers knowledge from labeled source domain to the unlabeled target domain Lu et al. (2025), categorized into adversarial Ganin et al. (2016), statistical Gretton et al. (2012), and self-supervised He et al. (2023). For time series in UDA, methods similarly fall into adversarial, statistical, and contrastive learning, but emphasize temporal properties such as causal structures, frequency, and periodicity Li et al. (2024); Liu & Xue (2021); Pérez-Carrasco et al. (2019); Wilson et al. (2020). Representative works include ACON Liu et al. (2024), EDEN Liu et al. (2026), GLA-DA Tu et al. (2024) in adversarial,AdvSKM Liu & Xue (2021) and RAINCOAT He et al. (2023) in statistical and CLUDA Ozyurt et al. (2022), CoTMix Jia et al. (2025), ConFGD Guo et al. (2025) in contrastive.

6.2 Maximum Mean Discrepancy

Maximum Mean Discrepancy (MMD) is a widely used distribution distance metric in domain adaptation Gretton et al. (2012). 1) Among marginal distribution alignment methods, first-order moment matching aligns mean embeddings in RKHS Long et al. (2015), second-order matching aligns covariance Sun & Saenko (2016), and higher-order matching Chen et al. (2020) aligns third-order and above central moments. 2) For joint distribution alignment JMMD Long et al. (2017), which measures the discrepancy of uncentered feature-label covariances across domains and unifies CMMD and WCMMD. DCAN Ge et al. (2020) minimizes CMMD to align conditional distributions, LMMD Zhu et al. (2020) weights samples to capture sub-domain structures, and WCMMD Huang et al. (2026); Lin et al. (2024) introduces class priors for class imbalance. JMMD-HSIC Wang et al. (2026) mitigates JMMD's discriminability degradation via HSIC with a designed label kernel, yet MMD-based methods remain unsuited to time series, where heuristic-bandwidth first-order matching enlarges intra-class dispersion and destroys sub-modal structures, motivating META.

6.3 Meta-Learning

Meta learning extracts transferable prior knowledge from related tasks Lu et al. (2026); Zhang et al. (2025). A key application is hyperparameter optimization via bi-level optimization: the inner loop trains a model with candidate hyperparameters, and the outer loop updates the meta-parameters that generate them Przewięźlikowski et al. (2024); Lin et al. (2025a). Recent works accelerate this through task similarity, stochastic hypergradients, or surrogate models Wei et al. (2021); Gao et al. (2026); Baik et al. (2023). We adopt meta learning to optimize MMD Gaussian kernel bandwidths, adaptively distinguishing cross-sub-mode from within-sub-mode pairs for fine-grained alignment.

AI USE STATEMENT

In this work, we used generative AI tools for polishing and editing the manuscript (a task with recommended disclosure). We have not used generative AI tools for generating synthetic data, proposing hypotheses or methodology, implementing code, constructing theoretical models, mathematical statements, or proofs, or performing translation, data cleaning, qualitative analysis, or result interpretation, and none of the required disclosure tasks are applicable to this work. We have reviewed all AI-assisted work. Specifically, the polished text was checked to ensure it faithfully preserves the original meaning and technical claims. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

See the ICLR 2027 AI Policy for Authors for more details. This statement should not be more than 1 page.

ETHICS STATEMENT

This work focuses on developing machine learning methods for cross-domain human activity recognition using publicly available datasets. We do not involve any human subjects or collect new data. All datasets used are de-identified and publicly available for research purposes. There are no conflicts of interest. All authors have followed ethical research practices, and we are committed to releasing code and models for reproducibility.

REFERENCES

Sungyong Baik, Myungsub Choi, Janghoon Choi, Heewon Kim, and Kyoung Mu Lee. Learning to learn task-adaptive hyperparameters for few-shot learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1441–1454, 2023.

Chao Chen, Zhihang Fu, Zhihong Chen, Sheng Jin, Zhaowei Cheng, Xinyu Jin, and Xian-Sheng Hua. Homm: Higher-order moment matching for unsupervised domain adaptation. In *Proceedings of the AAAI conference on artificial intelligence*, pp. 3422–3429, 2020.

Hohyun Cho, Minkyu Ahn, Sangtae Ahn, Moonyoung Kwon, and Sung Jun. Eeg datasets for motor imagery brain–computer interface. *GigaScience*, 2017.

Wonjeong Choi, Do-Yeon Kim, Jungwuk Park, Jungmoon Lee, Younghyun Park, Dong-Jun Han, and Jaekyun Moon. Adaptive energy alignment for accelerating test-time adaptation. In *International Conference on Learning Representations*, volume 2025, pp. 21197–21219, 2025.

Zahra Zamanzadeh Darban, Yiyuan Yang, Geoffrey I Webb, Charu C Aggarwal, Qingsong Wen, Shirui Pan, and Mahsa Salehi. Dacad: Domain adaptation contrastive learning for anomaly detection in multivariate time series. *IEEE Transactions on Knowledge and Data Engineering*, 2025.

Jean-Christophe Gagnon-Audet, Kartik Ahuja, Mohammad-Javad Darvishi-Bayazi, Pooneh Mousavi, Guillaume Dumas, and Irina Rish. WOODS: Benchmarks for Out-of-Distribution Generalization in Time Series. *arXiv e-prints*, pp. arXiv:2203.09978, 2022.

Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario March, and Victor Lempitsky. Domain-adversarial training of neural networks. *Journal of machine learning research*, pp. 1–35, 2016.

Ang Gao, Changshuo Zhang, Xiao Zhang, Deyang Li, Minjun Zhao, Fangchao Liu, and Xinyu Zhang. Process in-context learning: Enhancing mathematical reasoning via dynamic demonstration insertion. *arXiv preprint arXiv:2601.11979*, 2026.

Daniel Garcia-Gonzalez, Daniel Rivero, Enrique Fernandez-Blanco, and Miguel R Luaces. A public domain dataset for real-life human activity recognition using smartphone sensors. *Sensors*, pp. 2200, 2020.

Pengfei Ge, Chuan-Xian Ren, Dao-Qing Dai, and Hong Yan. Domain adaptation and image classification via deep conditional adaptation network. *arXiv preprint arXiv:2006.07776*, 2020.

Ary Goldberger, Luís Amaral, L. Glass, Shlomo Havlin, J. Hausdorg, Plamen Ivanov, R. Mark, J. Mietus, G. Moody, Chung-Kang Peng, H. Stanley, and Physiotoolkit Physiobank. Components of a new research resource for complex physiologic signals. *PhysioNet*, 2000.

Arthur Gretton, Karsten M Borgwardt, Malte J Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *The journal of machine learning research*, pp. 723–773, 2012.

Haoren Guo, Haiyue Zhu, Jiahui Wang, Prahlad Vadakkepat, Weng Khuen Ho, and Tong Heng Lee. Enhancing multivariate time-series domain adaptation via contrastive frequency graph discovery and language-guided adversary alignment. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 16951–16959, 2025.

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*, pp. 12746–12774. PMLR, 2023.

Kaihui Huang, Runqing Wu, JinHui Sheng, Hanyi Zhang, Ling Ge, Jinyu Guo, and Fei Ye. Learning dynamic representations via an optimally-weighted maximum mean discrepancy optimization framework for continual learning. *Knowledge-Based Systems*, pp. 115419, 2026.

Zidi Jia, Lei Ren, and Yang Tang. A contrastive representation domain adaptation method for industrial time-series cross-domain prediction. *IEEE Transactions on Industrial Informatics*, pp. 3236–3245, 2025.

Panagiotis Koromilas, Giorgos Bouritsas, Theodoros Giannakopoulos, Mihalis Nicolaou, and Yannis Panagakis. Bridging mini-batch and asymptotic analysis in contrastive learning: From infonce to kernel-based losses. *arXiv preprint arXiv:2405.18045*, 2024.

Jennifer R Kwapisz, Gary M Weiss, and Samuel A Moore. Activity recognition using cell phone accelerometers. *ACM SigKDD Explorations Newsletter*, pp. 74–82, 2011.

Hualing Li, Yaodan Wang, and Yan Qiang. A semi-supervised domain adaptive medical image segmentation method based on dual-level multi-scale alignment. *Scientific Reports*, 15(1):8784, 2025.

Zijian Li, Ruichu Cai, Jiawei Chen, Yuguang Yan, Wei Chen, Keli Zhang, and Junjian Ye. Time-series domain adaptation via sparse associative structure alignment: Learning invariance and variance. *Neural Networks*, 2024.

Hsiau-Wen Lin, Yihjia Tsai, Hwei Jen Lin, Chen-Hsiang Yu, and Meng-Hsing Liu. Unsupervised domain adaptation with deep network based on discriminative class-wise mmd. *AIMS Math*, 9 (3):6628–6647, 2024.

Hsiau-Wen Lin, Trang-Thi Ho, Ching-Ting Tu, Hwei-Jen Lin, and Chen-Hsiang Yu. Meta learning-based optimization of unsupervised domain adaptation deep networks. *Mathematics*, pp. 226, 2025a.

Yuena Lin, Haichun Cai, Jun-Yi Hang, Haobo Wang, Zhen Yang, and Gengyu Lyu. Mitigating local cohesion and global sparseness in graph contrastive learning with fuzzy boundaries. In *Forty-second International Conference on Machine Learning*, 2025b.

Mingyang Liu, Xinyang Chen, Yang Shu, Xiucheng Li, Weili Guan, and Liqiang Nie. Boosting transferability and discriminability for time series domain adaptation. *Advances in Neural Information Processing Systems*, pp. 100402–100427, 2024.

Mingyang Liu, Xinyang Chen, Xiucheng Li, Weili Guan, and Liqiang Nie. Breakthrough sensor-limited single view: Towards implicit temporal dynamics for time series domain adaptation. *Advances in Neural Information Processing Systems*, pp. 93384–93408, 2026.

Qiao Liu and Hui Xue. Adversarial spectral kernel matching for unsupervised time series domain adaptation. In *International Joint Conference on Artificial Intelligence*, pp. 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*, pp. 97–105. PMLR, 2015.

Mingsheng Long, Han Zhu, Jianmin Wang, and Michael I Jordan. Deep transfer learning with joint adaptation networks. In *International conference on machine learning*, pp. 2208–2217, 2017.

Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. Conditional adversarial domain adaptation. *Advances in neural information processing systems*, 2018.

Menglong Lu, Zejiang He, Yaohui Guo, Shanshan Liu, Jingyuan Huang, Yunxiang Zhao, Zhiliang Tian, Xiaoran Zhao, Chengcheng Shao, Lin Deng, et al. Llm-driven adversarial example synthesis for emerging topic rumor detection on social media. *IEEE Transactions on Knowledge and Data Engineering*, 2026.

Wang Lu, Jindong Wang, Xinwei Sun, Yiqiang Chen, and Xing Xie. Out-of-Distribution Representation Learning for Time Series Classification. *arXiv e-prints*, pp. arXiv:2209.07027, 2022.

Yuwu Lu, Xue Hu, Waikeung Wong, and Haoyu Huang. Collaborative semantic consistency alignment for blended-target domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 19170–19179, 2025.

Isao Matsui, Ayumi Matsumoto, Atsuhiro Imai, Hiroki Okushima, Hirohiko Niioka, Masatoshi Abe, Natsune Tamai, Hajime Nagasu, Eiichiro Kanda, Eiichiro Uchino, et al. Domain-adaptive semi-supervised learning for efficient rare pathological lesion detection with minimal annotation. *npj Digital Medicine*, 8(1):778, 2025.

Dragoslav S Mitrinovic, Josip Pecaric, and Arlington M Fink. *Classical and new inequalities in analysis*. Springer Science & Business Media, 2013.

Yilmazcan Ozyurt, Stefan Feuerriegel, and Ce Zhang. Contrastive learning for unsupervised domain adaptation of time series. *arXiv preprint arXiv:2206.06243*, 2022.

Manuel Pérez-Carrasco, Guillermo Cabrera-Vives, Pavlos Protopapas, Nicolás Astorga, and Marouan Belhaj. Matching Embeddings for Domain Adaptation. *arXiv e-prints*, 2019.

Marcin Przewięźlikowski, Przemysław Przybysz, Jacek Tabor, Maciej Zięba, and Przemysław Spurek. Hypermaml: Few-shot adaptation of deep models with hypernetworks. *Neurocomputing*, pp. 128179, 2024.

Mohammad Mahfujur Rahman, Clinton Fookes, Mahsa Baktashmotlagh, and Sridha Sridharan. On minimum discrepancy estimation for deep domain adaptation. In *Domain adaptation for visual understanding*, pp. 81–94. Springer, 2020.

J Michael Steele. *The Cauchy-Schwarz master class: an introduction to the art of mathematical inequalities*. Cambridge University Press, 2004.

Allan Stisen, Henrik Blunck, Sourav Bhattacharya, Thor Siiger Prentow, Mikkel Baun Kjærgaard, Anind Dey, Tobias Sonne, and Mads Møller Jensen. Smart devices are different: Assessing and mitigatingmobile sensing heterogeneities for activity recognition. In *Proceedings of the 13th ACM conference on embedded networked sensor systems*, pp. 127–140, 2015.

Baochen Sun and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. In *European conference on computer vision*, pp. 443–450. Springer, 2016.

Rahul Thapa, Magnus Ruud Kjaer, Bryan He, Ian Covert, Hyatt Moore IV, Umaer Hanif, Gauri Ganjoo, M Brandon Westover, Poul Jennum, Andreas Brink-Kjaer, et al. A multimodal sleep foundation model for disease prediction. *Nature Medicine*, pp. 1–11, 2026.

Gang Tu, Dan Li, Bingxin Lin, Zibin Zheng, and See-Kiong Ng. Gla-da: Global-local alignment domain adaptation for multivariate time series. In *International Conference on Database Systems for Advanced Applications*, pp. 476–492, 2024.

Wei Wang, Haifeng Xia, Chao Huang, Zhengming Ding, Cong Wang, Haojie Li, and Xiaochun Cao. Rethinking joint maximum mean discrepancy for visual domain adaptation. *Advances in Neural Information Processing Systems*, pp. 151296–151326, 2026.

Ying Wei, Peilin Zhao, and Junzhou Huang. Meta-learning hyperparameter performance prediction with neural processes. In *International Conference on Machine Learning*, pp. 11058–11067. PMLR, 2021.

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*, pp. 1768–1778, 2020.

Garrett Wilson, Janardhan Rao Doppa, and Diane J Cook. Calda: Improving multi-source time series domain adaptation with contrastive adversarial learning. *IEEE transactions on pattern analysis and machine intelligence*, pp. 14208–14221, 2023.

Shiguang Wu, Yaqing Wang, Yatao Bian, and Quanming Yao. Learning to learn with contrastive meta-objective. *Advances in Neural Information Processing Systems*, pp. 33275–33305, 2026.

Liyi Zhang, Jake C Snell, and Thomas L Griffiths. Amortized bayesian meta-learning for low-rank adaptation of large language models. In *Proceedings of the 2nd Workshop on Uncertainty-Aware NLP (UncertaiNLP 2025)*, pp. 194–199, 2025.

Yongchun Zhu, Fuzhen Zhuang, Jindong Wang, Guolin Ke, Jingwu Chen, Jiang Bian, Hui Xiong, and Qing He. Deep subdomain adaptation network for image classification. *IEEE transactions on neural networks and learning systems*, pp. 1713–1722, 2020.

A JMMD INDUCES INTRA-CLASS DISPERSION EXPANSION

The class-wise form of JMMD constrains only the class-conditional mean of the embeddings. Consequently, any mean-preserving perturbation lies in the null space of the objective: it changes neither the class mean nor the JMMD loss, and is therefore not penalized. We show that if other optimization forces scale such a drift beyond an explicit threshold, the intra-class covariance strictly increases, and so does V_{in}.

Assumption A.1. *For each class c, the JMMD loss depends on the class-c embeddings only through their class-conditional mean. Equivalently, for any perturbation $\Delta_{c,0} \in \mathbb{R}^{n_c \times D}$ with $\Delta_{c,0}^\top \mathbf{1}_{n_c} = 0$, the class mean and the JMMD loss are invariant under $\mathbf{Z}_c \mapsto \mathbf{Z}_c + t_c \Delta_{c,0}$ for all $t_c \in \mathbb{R}$.*

Lemma A.1 (JMMD Null-Space Drift Increases V_{in})**.** *Under Assumption A.1, let $\mathbf{Z}_c$ be the centered feature matrix, $\mathbf{Z}_c^\top \mathbf{1}_{n_c} = 0$, and $\mathbf{\Sigma}_c = n_c^{-1} \mathbf{Z}_c^\top \mathbf{Z}_c$. Suppose optimization produces a nonzero mean-preserving drift $\Delta_{c,0} \neq 0$ with $\Delta_{c,0}^\top \mathbf{1}_{n_c} = 0$, and set $\mathbf{Z}_c(t_c) = \mathbf{Z}_c + t_c \Delta_{c,0}$. Define $q_c(t) := \|\mathbf{\Sigma}_c(t)\|_F^2 - \|\mathbf{\Sigma}_c\|_F^2$ and $T_c := \max\{|r| : q_c(r) = 0\}$. Here T_c is the threshold such that for all $|t_c| > T_c$, $\|\mathbf{\Sigma}_c(t_c)\|_F > \|\mathbf{\Sigma}_c\|_F$. If $|t_c| > T_c$ holds for every class c, then*

$$V_{in}' = \sum_{c=1}^{C} \alpha_c \|\mathbf{\Sigma}_c(t_c)\|_F > \sum_{c=1}^{C} \alpha_c \|\mathbf{\Sigma}_c\|_F = V_{in}. \tag{12}$$

Proof. Since $\mathbf{Z}_c$ is centered and $\Delta_{c,0}$ is mean-preserving, the perturbed features remain centered: $\mathbf{1}_{n_c}^\top \mathbf{Z}_c(t_c) = \mathbf{1}_{n_c}^\top \mathbf{Z}_c + t_c \mathbf{1}_{n_c}^\top \Delta_{c,0} = 0$. Hence $\mathbf{\Sigma}_c(t_c) = \frac{1}{n_c} \mathbf{Z}_c(t_c)^\top \mathbf{Z}_c(t_c)$ is the covariance matrix.

$$\begin{aligned} \mathbf{\Sigma}_c(t_c) &= \frac{1}{n_c} \big(\mathbf{Z}_c + t_c \Delta_{c,0}\big)^\top \big(\mathbf{Z}_c + t_c \Delta_{c,0}\big) \\ &= \frac{1}{n_c} \mathbf{Z}_c^\top \mathbf{Z}_c + \frac{t_c}{n_c} \big(\mathbf{Z}_c^\top \Delta_{c,0} + \Delta_{c,0}^\top \mathbf{Z}_c\big) + \frac{t_c^2}{n_c} \Delta_{c,0}^\top \Delta_{c,0} \\ &= \mathbf{\Sigma}_c + \frac{t_c}{n_c} A_c + \frac{t_c^2}{n_c} B_c, \end{aligned} \tag{13}$$

where $A_c = \mathbf{Z}_c^\top \Delta_{c,0} + \Delta_{c,0}^\top \mathbf{Z}_c$ and $B_c = \Delta_{c,0}^\top \Delta_{c,0}$ are both symmetric. Expanding the squared Frobenius norm via $\|X\|_F^2 = \operatorname{tr}(X^\top X)$ we have

$$\begin{aligned} \|\mathbf{\Sigma}_c(t_c)\|_F^2 &= \operatorname{tr}\Big[\Big(\mathbf{\Sigma}_c + \tfrac{t_c}{n_c} A_c + \tfrac{t_c^2}{n_c} B_c\Big)^\top \Big(\mathbf{\Sigma}_c + \tfrac{t_c}{n_c} A_c + \tfrac{t_c^2}{n_c} B_c\Big)\Big] \\ &= \|\mathbf{\Sigma}_c\|_F^2 + \frac{2t_c}{n_c} \operatorname{tr}(\mathbf{\Sigma}_c A_c) + \frac{2t_c^2}{n_c} \operatorname{tr}(\mathbf{\Sigma}_c B_c) + \frac{t_c^2}{n_c^2} \|A_c\|_F^2 + \frac{2t_c^3}{n_c^2} \operatorname{tr}(A_c B_c) + \frac{t_c^4}{n_c^2} \|B_c\|_F^2. \end{aligned} \tag{14}$$

Since $\Delta_{c,0} \neq 0$, B_c is nonzero positive semidefinite, the leading coefficient of (14) satisfies $\frac{1}{n_c^2} \|B_c\|_F^2 > 0$. Therefore $q_c(t) = \|\mathbf{\Sigma}_c(t)\|_F^2 - \|\mathbf{\Sigma}_c\|_F^2$ is a quartic polynomial with $q_c(0) = 0$ and a strictly positive leading coefficient. Hence its real roots are finite and nonempty, so $T_c = \max\{|r| : q_c(r) = 0\}$ is well-defined. A positive-leading quartic is strictly positive outside its extreme roots, so $q_c(t) > 0$ for all $|t| > T_c$. Since the Frobenius norm is nonnegative,

$$|t_c| > T_c \implies q_c(t_c) > 0 \iff \|\mathbf{\Sigma}_c(t_c)\|_F^2 > \|\mathbf{\Sigma}_c\|_F^2 \iff \|\mathbf{\Sigma}_c(t_c)\|_F > \|\mathbf{\Sigma}_c\|_F. \tag{15}$$

Finally, if $|t_c| > T_c$ for every class c, then (15) holds class-wise, and since $\alpha_c > 0$ the strict inequality is preserved under weighted summation:

$$V_{in}' = \sum_{c=1}^{C} \alpha_c \|\mathbf{\Sigma}_c(t_c)\|_F > \sum_{c=1}^{C} \alpha_c \|\mathbf{\Sigma}_c\|_F = V_{in}. \tag{16}$$

Therefore the global intra-class dispersion strictly increases. □

B TIME-FREQUENCY CONTRASTIVE LEARNING COMPRESSES INTRA-CLASS DISPERSION

Given B_s source samples, each with a time and frequency-domain view, let $\mathcal{I} = \{1, \ldots, 2B_s\}$ index all views. For anchor i, $\mathcal{P}(i) = \{p \in \mathcal{I} \mid y_p = y_i, p \neq i\} \neq \varnothing$. Let $\mathcal{S}_c$ be the views of class c,

$|\mathcal{S}_c| = n_c$. The projected features $\mathbf{z}_i$ are L2-normalized to $\tilde{\mathbf{z}}_i \in \mathbb{S}^{D-1}$. With temperature $\tau > 0$, define $s_{ij} := \tilde{\mathbf{z}}_i^\top \tilde{\mathbf{z}}_j \in [-1, 1]$.

Assumption B.1. *The optimization of $\mathcal{L}_{\mathrm{SupCon}}$ drives the alignment term close to its lower bound, so that every positive pair is nearly coincident, i.e., $\|\tilde{\mathbf{z}}_i - \tilde{\mathbf{z}}_p\|_2 \le \epsilon$ for all $i \in \mathcal{I}$, $p \in \mathcal{P}(i)$, for some small $\epsilon > 0$.*

Theorem B.1. *Suppose $\mathcal{L}_{SupCon}$ attains its global optimum and Assumption B.1 holds. Then: (i) for each class $c \in \mathcal{C}$, all normalized features of class c collapse to the same class center: $\tilde{\mathbf{z}}_i = \boldsymbol{\mu}_c$ for all $i \in \mathcal{S}_c$, with $\|\boldsymbol{\mu}_c\|_2 = 1$; (ii) the intra-class covariance matrix vanishes, $\boldsymbol{\Sigma}_c = \frac{1}{|\mathcal{S}_c|}\sum_{i\in\mathcal{S}_c}(\tilde{\mathbf{z}}_i - \boldsymbol{\mu}_c)(\tilde{\mathbf{z}}_i - \boldsymbol{\mu}_c)^\top = \mathbf{0}$, and consequently*

$$\|\boldsymbol{\Sigma}_c\|_F = 0, \qquad V_{in} = \sum_{c=1}^{C} \alpha_c \|\boldsymbol{\Sigma}_c\|_F = 0; \tag{17}$$

(iii) if the batch is class-balanced ($n_c = m$ for all c) and $C \le D + 1$, the class centers minimizing the uniformity term form a regular simplex on $\mathbb{S}^{D-1}$: $\boldsymbol{\mu}_c^\top \boldsymbol{\mu}_{c'} = -\frac{1}{C-1}$ and $\sum_{c=1}^{C} \boldsymbol{\mu}_c = \mathbf{0}$.

Proof. Starting from (3), the supervised contrastive loss decomposes as

$$\begin{aligned}
\mathcal{L}_{\mathrm{SupCon}} &= -\sum_{i\in\mathcal{I}} \frac{1}{|\mathcal{P}(i)|} \sum_{p\in\mathcal{P}(i)} \left[\frac{s_{ip}}{\tau} - \log \sum_{j\neq i} \exp\Big(\frac{s_{ij}}{\tau}\Big) \right] \\
&= \underbrace{-\sum_{i\in\mathcal{I}} \frac{1}{|\mathcal{P}(i)|} \sum_{p\in\mathcal{P}(i)} \frac{s_{ip}}{\tau}}_{\mathcal{L}_{\text{align}}} + \underbrace{\sum_{i\in\mathcal{I}} \frac{1}{|\mathcal{P}(i)|} \sum_{p\in\mathcal{P}(i)} \log \sum_{j\neq i} \exp\Big(\frac{s_{ij}}{\tau}\Big)}_{\mathcal{L}_{\text{unif}}} \\
&\ge -\frac{1}{\tau} \sum_{i\in\mathcal{I}} \frac{1}{|\mathcal{P}(i)|} \sum_{p\in\mathcal{P}(i)} \|\tilde{\mathbf{z}}_i\|_2 \|\tilde{\mathbf{z}}_p\|_2 + \mathcal{L}_{\text{unif}} \\
&= -\frac{1}{\tau} \sum_{i\in\mathcal{I}} \frac{1}{|\mathcal{P}(i)|} \sum_{p\in\mathcal{P}(i)} 1 + \mathcal{L}_{\text{unif}} \\
&= -\frac{|\mathcal{I}|}{\tau} + \mathcal{L}_{\text{unif}}.
\end{aligned} \tag{18}$$

Here $\mathcal{L}_{\text{align}}$ and $\mathcal{L}_{\text{unif}}$ denote the alignment and uniformity terms, respectively. (18) holds iff $s_{ip} = 1$ for all $i \in \mathcal{I}$, $p \in \mathcal{P}(i)$. For unit vectors, this is the equality case of Cauchy–Schwarz Steele (2004), forcing $\tilde{\mathbf{z}}_i = \tilde{\mathbf{z}}_p$. Since $\mathcal{P}(i) \neq \varnothing$ and same-class relation is transitive, all normalized features in each class c coincide: there exists $\boldsymbol{\mu}_c \in \mathbb{S}^{D-1}$ such that $\tilde{\mathbf{z}}_i = \boldsymbol{\mu}_c$ for all $i \in \mathcal{S}_c$. Equivalently, the empirical class-conditional distribution degenerates to the Dirac measure $\delta_{\boldsymbol{\mu}_c}$. By Assumption B.1, the global optimum attains equality, proving claim (i). At collapse, the intra-class covariance evaluates to

$$\boldsymbol{\Sigma}_c = \frac{1}{|\mathcal{S}_c|} \sum_{i\in\mathcal{S}_c} (\tilde{\mathbf{z}}_i - \boldsymbol{\mu}_c)(\tilde{\mathbf{z}}_i - \boldsymbol{\mu}_c)^\top = \frac{1}{|\mathcal{S}_c|} \sum_{i\in\mathcal{S}_c} (\boldsymbol{\mu}_c - \boldsymbol{\mu}_c)(\boldsymbol{\mu}_c - \boldsymbol{\mu}_c)^\top = \mathbf{0}, \tag{19}$$

Thus $\|\mathbf{\Sigma}_c\|_F = 0$ and $V_{in} = 0$, proving (ii). For (iii), at collapse the uniformity term depends only on the class centers. Writing$s_{cc'} := \boldsymbol{\mu}_c^\top \boldsymbol{\mu}_{c'}$ and assuming a class-balanced batch $n_c = m$,

$$\begin{aligned}
\mathcal{L}_{\text{unif}} &= \sum_{i\in\mathcal{I}} \frac{1}{|\mathcal{P}(i)|} \sum_{p\in\mathcal{P}(i)} \log\Big[\sum_{p'\in\mathcal{P}(i)} e^{1/\tau} + \sum_{c'\neq c} \sum_{j\in\mathcal{S}_{c'}} e^{s_{cc'}/\tau}\Big] \\
&= \sum_{c=1}^{C} m \log\Big((m-1)e^{1/\tau} + m\sum_{c'\neq c} e^{s_{cc'}/\tau}\Big) \\
&\geq m\sum_{c=1}^{C} \log\Big((m-1)e^{1/\tau} + m(C-1)\, e^{\bar{s}_c/\tau}\Big), \qquad \bar{s}_c := \frac{1}{C-1}\sum_{c'\neq c} s_{cc'} \\
&\geq mC \log\Big((m-1)e^{1/\tau} + m(C-1)\, e^{\bar{s}/\tau}\Big), \qquad \bar{s} := \frac{1}{C}\sum_{c=1}^{C} \bar{s}_c \\
&\geq mC \log\Big((m-1)e^{1/\tau} + m(C-1)\, e^{-\frac{1}{\tau(C-1)}}\Big).
\end{aligned} \tag{20}$$

Here the first inequality applies Jensen Mitrinovic et al. (2013) to exp, the second to the convex function $g(s) = \log\big((m-1)e^{1/\tau} + m(C-1)e^{s/\tau}\big)$, and the third uses the Gram constraint Mitrinovic et al. (2013) and the monotonicity of g:

$$\begin{aligned}
\bar{s} &= \frac{1}{C(C-1)} \sum_{c\neq c'} s_{cc'} \\
&= \frac{1}{C(C-1)} \Big(\Big\| \sum_{c=1}^{C} \boldsymbol{\mu}_c \Big\|_2^2 - \sum_{c=1}^{C} \|\boldsymbol{\mu}_c\|_2^2 \Big) \\
&= \frac{1}{C(C-1)} \Big(\Big\| \sum_{c=1}^{C} \boldsymbol{\mu}_c \Big\|_2^2 - C \Big) \\
&\geq -\frac{1}{C-1}.
\end{aligned} \tag{21}$$

(20) to (21) requires $s_{cc'} = s^*$ for all $c \neq c'$ and $\|\sum_c \boldsymbol{\mu}_c\|_2 = 0$, yielding $s^* = -1/(C-1)$. Hence the minimizing centers form a regular simplex, $\boldsymbol{\mu}_c^\top \boldsymbol{\mu}_{c'} = -1/(C-1)$ with $\sum_c \boldsymbol{\mu}_c = \mathbf{0}$, feasible on $\mathbb{S}^{D-1}$ iff $C \leq D+1$. □

Under Assumption B.1, if $\mathcal{L}_{\text{SupCon}}$ attains its global optimum, alignment dominates: all normalized features within each class collapse to a single center $\boldsymbol{\mu}_c$, so $\mathbf{\Sigma}_c = \mathbf{0}$ and $V_{in} = 0$. The uniformity term then arranges the class centers $\{\boldsymbol{\mu}_c\}$ as evenly as possible on $\mathbb{S}^{D-1}$. When $C \leq D+1$, the minimizer is a regular simplex with $\boldsymbol{\mu}_c^\top \boldsymbol{\mu}_{c'} = -1/(C-1)$ for $c \neq c'$, not the orthogonal configuration $\boldsymbol{\mu}_c^\top \boldsymbol{\mu}_{c'} = 0$, because $e^{-1/(\tau(C-1))} < 1 = e^0$: the simplex attains a strictly smaller uniformity value. Orthogonal centers arise only under a non-negative cross-inner-product constraint or as an approximation when $C \approx D$. For $C > D+1$, the optimum is a Grassmannian frame minimizing the maximum pairwise inner product, and the simplex lower bound in (20) remains valid but is not attained. This arrangement affects inter-class separation, not intra-class dispersion: V_{in} is already zero from the within-class collapse, while $\mathcal{L}_{\text{align}}$ guarantees intra-class compactness.

C Contrastive Stretching Enlarges Sub-Modal Dispersion

For sub-mode m of class c, let its $n_{c,m}$ samples have original centered features $u_i = z_i^o - \boldsymbol{\mu}_{c,m}^o$, $i = 1, \ldots, n_{c,m}$, and covariance$\mathbf{\Sigma}_{c,m} = \frac{1}{n_{c,m}} \sum_{i=1}^{n_{c,m}} u_i u_i^\top$. Let $\delta_i = z_i' - z_i^o$ be the displacement induced by contrastive learning, with sub-modal mean drift $\Delta\boldsymbol{\mu}_{c,m} := \frac{1}{n_{c,m}} \sum_{i=1}^{n_{c,m}} \delta_i$ and centered displacement $r_i := \delta_i - \Delta\boldsymbol{\mu}_{c,m}$, so that $\sum_i r_i = \sum_i u_i = 0$.

Assumption C.1. *For a sub-mode (c, m) with $\mathbf{\Sigma}_{c,m} \neq \mathbf{0}$, we say that the proportional stretching condition holds if the centered displacement admits the decomposition $r_i = \lambda_{c,m} u_i + \epsilon_i$, where $\lambda_{c,m} > 0$ and the residual ϵ_i satisfies $\sum_{i=1}^{n_{c,m}} \epsilon_i = 0$ and $\sum_{i=1}^{n_{c,m}} u_i \epsilon_i^\top = \mathbf{0}$.*

Equivalently, $\sum_i u_i r_i^\top = \lambda_{c,m} \sum_i u_i u_i^\top$, so the cross-moment is a positive multiple of the within-sub-mode second-moment matrix. Hence the non-uniform pull is an isotropic radial stretch of the sub-mode, the only substantive content of Assumption C.1.

Proposition C.1. *Under Assumption C.1, the displaced sub-modal covariance satisfies* $\|\mathbf{\Sigma}'_{c,m}\|_F \geq (1+\lambda_{c,m})^2 \|\mathbf{\Sigma}_{c,m}\|_F > \|\mathbf{\Sigma}_{c,m}\|_F$ *whenever* $\mathbf{\Sigma}_{c,m} \neq \mathbf{0}$. *If Assumption C.1 holds for every sub-mode, then, with* $\beta_{c,m} > 0$,

$$V'_{sub} = \sum_{c=1}^{C} \sum_{m=1}^{M_c} \beta_{c,m} \|\mathbf{\Sigma}'_{c,m}\|_F > \sum_{c=1}^{C} \sum_{m=1}^{M_c} \beta_{c,m} \|\mathbf{\Sigma}_{c,m}\|_F = V_{sub}. \tag{22}$$

Proof. The displaced sub-modal mean is

$$\begin{aligned} \boldsymbol{\mu}'_{c,m} &= \frac{1}{n_{c,m}} \sum_{i=1}^{n_{c,m}} z'_i = \boldsymbol{\mu}^o_{c,m} + \frac{1}{n_{c,m}} \sum_{i=1}^{n_{c,m}} \delta_i \\ &= \boldsymbol{\mu}^o_{c,m} + \Delta\boldsymbol{\mu}_{c,m} + \frac{\lambda_{c,m}}{n_{c,m}} \sum_{i=1}^{n_{c,m}} u_i + \frac{1}{n_{c,m}} \sum_{i=1}^{n_{c,m}} \epsilon_i \\ &= \boldsymbol{\mu}^o_{c,m} + \Delta\boldsymbol{\mu}_{c,m}, \end{aligned} \tag{23}$$

since $\sum_i u_i = \sum_i \epsilon_i = 0$, and the new centered feature is

$$\begin{aligned} u'_i &= z'_i - \boldsymbol{\mu}'_{c,m} \\ &= \big(\boldsymbol{\mu}^o_{c,m} + u_i + \Delta\boldsymbol{\mu}_{c,m} + \lambda_{c,m} u_i + \epsilon_i\big) - \big(\boldsymbol{\mu}^o_{c,m} + \Delta\boldsymbol{\mu}_{c,m}\big) \\ &= (1+\lambda_{c,m}) u_i + \epsilon_i. \end{aligned} \tag{24}$$

Hence, with $\sum_i u_i \epsilon_i^\top = \mathbf{0}$:

$$\begin{aligned} \mathbf{\Sigma}'_{c,m} &= \frac{1}{n_{c,m}} \sum_{i=1}^{n_{c,m}} u'_i u'^\top_i \\ &= \frac{1}{n_{c,m}} \sum_{i=1}^{n_{c,m}} \big((1+\lambda_{c,m}) u_i + \epsilon_i\big)\big((1+\lambda_{c,m}) u_i + \epsilon_i\big)^\top \\ &= (1+\lambda_{c,m})^2 \mathbf{\Sigma}_{c,m} + \frac{1+\lambda_{c,m}}{n_{c,m}} \sum_{i=1}^{n_{c,m}} \big(u_i \epsilon_i^\top + \epsilon_i u_i^\top\big) + \frac{1}{n_{c,m}} \sum_{i=1}^{n_{c,m}} \epsilon_i \epsilon_i^\top \\ &= (1+\lambda_{c,m})^2 \mathbf{\Sigma}_{c,m} + E_{c,m}, \end{aligned} \tag{25}$$

where $E_{c,m} = \frac{1}{n_{c,m}} \sum_{i=1}^{n_{c,m}} \epsilon_i \epsilon_i^\top \succeq \mathbf{0}$. Since both summands are positive semi-definite and $\mathrm{tr}(\mathbf{\Sigma}_{c,m} \epsilon_i \epsilon_i^\top) = \epsilon_i^\top \mathbf{\Sigma}_{c,m} \epsilon_i \geq 0$, we have

$$\begin{aligned} \|\mathbf{\Sigma}'_{c,m}\|_F^2 &= \big\|(1+\lambda_{c,m})^2 \mathbf{\Sigma}_{c,m} + E_{c,m}\big\|_F^2 \\ &= (1+\lambda_{c,m})^4 \|\mathbf{\Sigma}_{c,m}\|_F^2 + 2(1+\lambda_{c,m})^2 \,\mathrm{tr}\big(\mathbf{\Sigma}_{c,m} E_{c,m}\big) + \|E_{c,m}\|_F^2 \\ &= (1+\lambda_{c,m})^4 \|\mathbf{\Sigma}_{c,m}\|_F^2 + \frac{2(1+\lambda_{c,m})^2}{n_{c,m}} \sum_{i=1}^{n_{c,m}} \epsilon_i^\top \mathbf{\Sigma}_{c,m} \epsilon_i + \|E_{c,m}\|_F^2 \\ &\geq (1+\lambda_{c,m})^4 \|\mathbf{\Sigma}_{c,m}\|_F^2, \end{aligned} \tag{26}$$

and taking square roots yields

$$\|\mathbf{\Sigma}'_{c,m}\|_F \geq (1+\lambda_{c,m})^2 \,\|\mathbf{\Sigma}_{c,m}\|_F > \|\mathbf{\Sigma}_{c,m}\|_F, \qquad \lambda_{c,m} > 0,\ \mathbf{\Sigma}_{c,m} \neq \mathbf{0}. \tag{27}$$

Finally, since $\beta_{c,m} > 0$ and the inequality is strict for every non-degenerate sub-mode,

$$\begin{aligned} V'_{sub} &= \sum_{c=1}^{C}\sum_{m=1}^{M_c} \beta_{c,m}\|\boldsymbol{\Sigma}'_{c,m}\|_F \\ &\geq \sum_{c=1}^{C}\sum_{m=1}^{M_c} \beta_{c,m}(1+\lambda_{c,m})^2\|\boldsymbol{\Sigma}_{c,m}\|_F \\ &> \sum_{c=1}^{C}\sum_{m=1}^{M_c} \beta_{c,m}\|\boldsymbol{\Sigma}_{c,m}\|_F = V_{sub} \end{aligned} \tag{28}$$

□

D MetaMMD Bandwidth Differentiation Shrinks Sub-Modal Dispersion

Recall MetaMMD's bi-level structure: the inner loop updates the base model on the support set by $\theta' = \theta - \eta\nabla_\theta\mathcal{L}_b(\theta,\phi)$ with $\mathcal{L}_b(\theta,\phi) = \mathcal{L}_{\mathrm{MMD}}(\theta,\Lambda) + \lambda_{\mathrm{cls}}\mathcal{L}_{\mathrm{cls}}(\theta)$, where Λ collects pair-specific bandwidths σ_{ij} from the meta-model; the outer loop updates bandwidths on the query set by maximizing the domain-discrimination loss $\mathcal{L}_v$, driving features toward domain confusion. This meta-objective differentiates bandwidths to shrink sub-modal dispersion: weights of domain-aligning pairs are driven to their maxima while those of mis-aligning pairs go to zero (Lemma D.1), and the induced weights contract sub-modal dispersion (Lemma D.2). For pair (i,j) with distance $d_{ij} = \|z_i - z_j\|$, the Gaussian kernel $k_\sigma(z_i,z_j) = \exp(-d_{ij}^2/(2\sigma^2))$ induces the gradient weight

$$w_{ij} = w(\sigma_{ij};d_{ij}) = \frac{e^{-d_{ij}^2/(2\sigma_{ij}^2)}}{\sigma_{ij}^2}, \qquad w'(\sigma;d) = \frac{\partial w}{\partial\sigma} = \frac{w(\sigma;d)}{\sigma}\Big(\frac{d^2}{\sigma^2} - 2\Big), \tag{29}$$

so $w(\cdot;d)$ is unimodal: $w'(\sigma;d) = 0 \iff \sigma = \sigma^*(d) = d/\sqrt{2}$, the maximal value is $w(\sigma^*(d);d) = 2/(ed^2)$, and $w(\sigma;d) \to 0$ both as $\sigma \to 0$ and as $\sigma \to \infty$.

Assumption D.1. *Let $\tilde{g}_i$ be the outer gradient of $\mathcal{L}_v$ backpropagated to the embedding z_i, and let Δ_{ij} be the displacement that pair (i,j) induces on z_i in one inner step, with $q_{ij} := \tilde{g}_i^\top\Delta_{ij}$. There exists $q_0 > 0$ such that (i) $q_{ij} \geq q_0$ for cross-domain pairs within the same sub-mode, and (ii) $q_{ij} \leq -q_0$ for within-domain pairs and for cross-domain pairs across different sub-modes.*

Assume the confusion gradient at target z_i of sub-mode (c,m) is dominated by the direction to its matched source sub-mode: $\tilde{g}_i = \kappa_i(\boldsymbol{\mu}^s_{c,m} - z_i) + \xi_i$, $\kappa_i \geq \underline{\kappa} > 0$, $\|\xi_i\| \leq \xi$, and let $r_i = \|z_i - \boldsymbol{\mu}^s_{c,m}\|$. If the source sub-mode has radius $\leq \rho$, then for a matched cross-domain pair,

$$q_{ij} = \kappa_i(\boldsymbol{\mu}^s_{c,m} - z_i)^\top(z_{s,j} - z_i) + \xi_i^\top(z_{s,j} - z_i) \geq \underline{\kappa}\, r_i(r_i - \rho) - \xi(r_i + \rho) > 0, \tag{30}$$

and for a within-sub-mode target pair with $r_j < r_i$,

$$q_{ij} = -\kappa_i(\boldsymbol{\mu}^s_{c,m} - z_i)^\top(z_j - z_i) + \xi_i^\top(z_i - z_j) \leq -\underline{\kappa}\, r_i(r_i - r_j) + \xi(r_i + r_j) < 0. \tag{31}$$

The cross-sub-mode case follows whenever the confusion direction aligns better with the matched source sub-mode than with any unmatched one, the regime maintained by MetaMMD via suppressing cross-sub-mode pulls.

Lemma D.1. *Under Assumption D.1, along the outer-loop flow $\dot{\sigma}_{ij} = \beta\, d\mathcal{L}_v/d\sigma_{ij}$ (pair distances quasi-static per outer step): (i) each same-sub-mode cross-domain pair has w^s_{ij} increasing monotonically to its maximum $2/(ed^2_{ij})$, so for matched-pair distances in $[d_-, d_+]$, $w^s_{\min} = \frac{2}{ed_+^2} \leq w^s_{ij} \leq \frac{2}{ed_-^2} = w^s_{\max}$; (ii) each within-domain or cross-sub-mode pair has w_{ij} decreasing monotonically to 0, so for any $\bar{w} < w^s_{\min}/2$ there is a finite outer-training time after which $w_{ij} \leq \bar{w}$, giving $w^s_{\min} > 2\bar{w}$.*

Proof. One inner step displaces z_i by $-\eta\nabla_{z_i}\mathcal{L}_{\mathrm{MMD}}$, with pair (i,j) contributing $\eta c_{ij} w_{ij}\Delta_{ij}$, where $c_{ij} > 0$ is the pair normalization. The outer-objective sensitivity and weight dynamics form the

chain

$$\frac{d\mathcal{L}_v}{d\sigma_{ij}} = \tilde{g}_i^\top \frac{d(\eta c_{ij} w_{ij} \Delta_{ij})}{d\sigma_{ij}} = \eta c_{ij} q_{ij} w'(\sigma_{ij}) \implies \dot{w}_{ij} = \beta \eta c_{ij} q_{ij} [w'(\sigma_{ij})]^2. \tag{32}$$

Since $[w'(\sigma_{ij})]^2 \geq 0$, the sign of $\dot{w}_{ij}$ equals that of q_{ij}. For a same-sub-mode cross-domain pair, Assumption D.1(i) gives $\dot{w}^s_{ij} \geq 0$, and σ_{ij} converges monotonically to $\sigma^*(d_{ij}) = d_{ij}/\sqrt{2}$, so

$$w^s_{ij} \longrightarrow \frac{2}{ed^2_{ij}} \in \left[\frac{2}{ed^2_+}, \frac{2}{ed^2_-}\right] = [w^s_{\min}, w^s_{\max}]. \tag{33}$$

For a within-domain or cross-sub-mode pair, Assumption D.1(ii) gives $\dot{w}_{ij} \leq 0$, and σ_{ij} is driven to 0 or $+\infty$, yielding $w_{ij} \to 0$. On the large-bandwidth branch, $w \sim \sigma^{-2}$ and $[w']^2 \sim 4w^3$, hence $\dot{w}_{ij} \leq -4\beta\eta c_{ij} q_0 w^3_{ij}$, which integrates to

$$w_{ij}(t) \leq \left(w_{ij}(0)^{-2} + 8\beta\eta c_{ij} q_0 t\right)^{-1/2}. \tag{34}$$

Thus for any $\bar{w} < w^s_{\min}/2$, there exists a finite time T such that $w_{ij}(t) \leq \bar{w}$ for all $t \geq T$, giving $2\bar{w} < w^s_{\min}$. □

Lemma D.2. *Let the weight configuration of Lemma D.1 hold: within a target sub-mode $w_{ij} = w_{ji} \leq \bar{w}$, matched cross-domain pairs satisfy $w^s_{\min} \leq w^s_{ij} \leq w^s_{\max}$ with $w^s_{\min} > 2\bar{w}$, and the source sub-mode satisfies $\frac{1}{n^s_{c,m}} \sum_j \|z_{s,j} - \boldsymbol{\mu}^s_{c,m}\|^2 \leq \rho^2$. The inner flow then satisfies, with $\gamma_{c,m} = \frac{4\eta}{n_{c,m}}(w^s_{\min} - 2\bar{w}) > 0$ and $M_{c,m}(t) = \frac{1}{n_{c,m}} \sum_i \|z_i(t) - \boldsymbol{\mu}^s_{c,m}\|^2$,*

$$V_{c,m}(t) \leq \left(e^{-\gamma_{c,m} t/2} \sqrt{M_{c,m}(0)} + \frac{w^s_{\max} \rho}{w^s_{\min} - 2\bar{w}}\right)^2. \tag{35}$$

If $\rho = 0$, then $V_{c,m}(t) \leq M_{c,m}(0) e^{-\gamma_{c,m} t} \to 0$, so $\|\boldsymbol{\Sigma}_{c,m}(t)\|_F \to 0$ for every sub-mode and $V_{sub}(t) \to 0$.

Proof. Since $\nabla_{z_i} k_{\sigma_{ij}}(z_i, z_j) = w_{ij}(z_j - z_i)$ and $w_{ij} = w_{ji}$, the inner flow is $\dot{z}_i = \frac{2\eta}{n^2} \sum_j w_{ij}(z_i - z_j) + \frac{2\eta}{n n_s} \sum_j w^s_{ij}(z_{s,j} - z_i)$, with $n = n_{c,m}$, $n_s = n^s_{c,m}$, $\boldsymbol{\mu}^s = \boldsymbol{\mu}^s_{c,m}$. We have

$$\begin{aligned}
\frac{dM_{c,m}}{dt} &= \frac{2}{n} \sum_i (z_i - \boldsymbol{\mu}^s)^\top \dot{z}_i \\
&= \frac{4\eta}{n^3} \sum_{i,j} w_{ij} (z_i - \boldsymbol{\mu}^s)^\top (z_i - z_j) - \frac{4\eta}{n^2 n_s} \sum_{i,j} w^s_{ij} (z_i - \boldsymbol{\mu}^s)^\top (z_i - z_{s,j}) \\
&\leq \frac{4\eta\bar{w}}{n^3} \cdot 2n^2 M_{c,m} - \frac{4\eta}{n^2 n_s} \left[w^s_{\min} n n_s M_{c,m} - w^s_{\max} n n_s \rho \sqrt{M_{c,m}}\right] \\
&= -\gamma_{c,m} M_{c,m} + \kappa \sqrt{M_{c,m}},
\end{aligned} \tag{36}$$

with $\kappa = \frac{4\eta w^s_{\max} \rho}{n}$. With $R(t) = \sqrt{M_{c,m}(t)}$, $\dot{R} \leq -\frac{\gamma_{c,m}}{2} R + \frac{\kappa}{2}$, so $V_{c,m} = \min_v \frac{1}{n} \sum_i \|z_i - v\|^2 \leq M_{c,m}$, give

$$V_{c,m}(t) \leq M_{c,m}(t) \leq \left(e^{-\gamma_{c,m} t/2} \sqrt{M_{c,m}(0)} + \frac{\kappa}{\gamma_{c,m}}\right)^2, \tag{37}$$

where $\frac{\kappa}{\gamma_{c,m}} = \frac{w^s_{\max} \rho}{w^s_{\min} - 2\bar{w}}$. If $\rho = 0$, then $V_{c,m}(t) \leq M_{c,m}(0) e^{-\gamma_{c,m} t} \to 0$, and since $\boldsymbol{\Sigma}_{c,m} \succeq \mathbf{0}$ implies $\|\boldsymbol{\Sigma}_{c,m}\|_F \leq \mathrm{tr}(\boldsymbol{\Sigma}_{c,m}) = V_{c,m}$, summing over sub-modes with $\beta_{c,m} > 0$ yields $V_{sub}(t) = \sum_{c,m} \beta_{c,m} \|\boldsymbol{\Sigma}_{c,m}(t)\|_F \leq \sum_{c,m} \beta_{c,m} M_{c,m}(0) e^{-\gamma_{c,m} t} \to 0$.

Since the total loss $\sum_{c,m} \mathcal{L}_{c,m}$ is separable across sub-modes and cross-sub-mode pairs carry vanishing weights by Lemma D.1(ii), each sub-mode contracts independently toward its matched source sub-mode without merging with other sub-modes, which preserves the multi-modal structure while shrinking V_{sub}. □

remark Assumption D.1 uses no sub-mode labels, and neither does any component of META; the assumption only requires that the sign of q_{ij} correlate with sub-mode membership, which is a geometric rather than a labeling property. This holds under the conditions that META itself creates: (i) after Time-Frequency Contrastive Learning, same-class features are compact and the multi-modal structure is preserved, so within each class the embedding distance to a target sample's own sub-mode is strictly smaller than to any other sub-mode; (ii) the domain discriminator D_ψ is trained only to separate source from target, so its confusion gradient at a target sample z_i points toward the region where source features are densest and nearest to z_i, i.e., toward the matched source sub-mode centroid $\boldsymbol{\mu}^s_{c,m}$ (cf. the decomposition $\tilde{g}_i = \kappa_i(\boldsymbol{\mu}^s_{c,m} - z_i) + \xi_i$ above); (iii) the cross term of the MMD loss couples only cross-domain pairs, so the outer-loop hypergradient $\nabla_\phi \mathcal{L}_v$ acts on exactly those pairs whose displacement direction is diagnostic of sub-mode membership, while within-domain pairs are driven to zero weight regardless. In other words, the domain discrimination loss and sub-mode discrimination are not required to coincide as objectives: the former supplies the direction of movement, and the latter is recovered from the alignment $q_{ij} = \tilde{g}_i^\top \Delta_{ij}$, whose sign is decided by the pair's relative geometry through. Moreover, since g_ϕ maps the embedding pair (z_i, z_j) to a bandwidth, identification is soft and continuous so the mechanism is robust to boundary samples and never needs a hard sub-mode partition. If the geometry assumption is violated, the sign of q_{ij} may flip for those pairs and MetaMMD gracefully degrades to standard heuristic-bandwidth MMD, since all bandwidths then receive comparable gradients.

E Details of META

E.1 Data Augmentation for Time-Frequency Contrastive Learning

E.1.1 Time-Domain Augmentation

Given a time series $x \in \mathbb{R}^{T \times d}$, we randomly select one of three strategies to produce the time-domain augmented view $\tilde{x}^t$:

- Random Cropping: Extract a subsequence of length T' ($T' < T$) starting from a random index.
- Random Time Masking: With probability 20% to 70%, mask out a contiguous time window covering 5% to 20% of the time axis.
- Additive Gaussian Noise: Add small Gaussian noise to the entire series.

E.1.2 Frequency-Domain Augmentation

Given x, we first apply Fast Fourier Transform (FFT) along the time dimension to obtain the frequency-domain tensor $X_f \in \mathbb{C}^{T \times d}$. We then apply a random mask to the magnitude spectrum $|X_f|$: a contiguous frequency interval covering 10% to 30% of the frequency axis is selected and its magnitude values are set to zero. The original phase spectrum $\angle X_f$ is fully preserved. The masked magnitude $|\tilde{X}_f|$ is recombined with $\angle X_f$, and the resulting complex spectrum is transformed back to the time domain via inverse FFT, yielding the frequency-domain augmented view $\tilde{x}^f$.

E.2 Temporal Extractor and Frequency Extractor

In the time series classification model, we train the temporal feature extractor E_t and the classifier H on the source domain with adversarial domain alignment. The frequency feature extractor E_f, while also trained on the source domain.

E.2.1 Temporal feature extraction.

Given an input time series $x_i \in \mathbb{R}^{T \times d}$, we extract temporal features using a 1D Convolutional Neural Network (CNN) as the temporal encoder E_t. The temporal feature is obtained as $z_i^t = E_t(x_i) \in \mathbb{R}^D$. The CNN architecture consists of three convolutional blocks, each followed by batch normalization, ReLU activation, and max-pooling.

E.2.2 FREQUENCY FEATURE EXTRACTION.

To capture domain-invariant spectral patterns, we transform the input signal into the frequency domain. The original time series x_i is first processed using a cosine smoothing: $\text{Smooth}(x_i) = x_i \cdot \left[0.5\left(1-\cos\left(\frac{2\pi t}{T-1}\right)\right)\right], \quad t = 0, 1, \ldots, T-1$, and then transformed into its frequency-domain representation X_f via the Fast Fourier Transform (FFT):

$$X_f[m] = \sum_{t=0}^{T-1} x[t] \cdot e^{-i \cdot \frac{2\pi m t}{T}}, \quad m \in [0, T-1].$$

Subsequently, the magnitude a_i and phase p_i of sample i are computed as $a_i[m] = |X_{f,i}[m]|/T, \quad p_i[m] = \text{atan2}(\text{Im}(X_{f,i}[m]), \text{Re}(X_{f,i}[m]))$, where m indexes the frequency bins. After applying convolution operations to X_f, we concatenate the magnitude and phase as the frequency feature $z_i^f = [a_i, p_i] \in \mathbb{R}^{2D_f}$: the magnitude component is inherently shift-invariant to temporal translations and thus stable across domains, while the phase component preserves complementary spectral structure. In addition, a lightweight projection head is applied to both z_i^t and z_i^f, mapping them into a shared embedding space of dimension D.

E.3 PSEUDO CODE

The overall training procedure of META is summarized in Algorithm 1.

Algorithm 1 META Training

Require: Source domain $\mathcal{D}_s$, target domain $\mathcal{D}_t$; learning rates η, β, γ; temperature τ; perturbation ϵ. Trained base model f_θ and meta-model g_ϕ.

Initialize θ, ϕ, ψ.
for epoch $= 1$ to E **do**
  Sample a batch from $\mathcal{D}_s$ and $\mathcal{D}_t$; split into support set $\mathcal{S}_\tau$ and query set $\mathcal{Q}_\tau$.
  for $k = 1$ to K **do**
    Extract $z_i^t = E_t(\tilde{x}_i^t), \quad z_i^f = [a_i, p_i]$. Project $\tilde{z}_i^\omega = P_\omega(z_i^\omega)/\|P_\omega(z_i^\omega)\|_2, \quad \omega \in \{t, f\}$.
    $\sigma_{ij} = g_\phi(z_i, z_j)$, and form Λ.
    Compute $\mathcal{L}_{\text{MMD}}(\theta, \Lambda)$, $\mathcal{L}_{SupCon}$, $\mathcal{L}_{TCon}$, $\mathcal{L}_{Cross}$, and $\mathcal{L}_{cls}(\theta)$.
    $\mathcal{L}_b(\theta, \phi) = \mathcal{L}_{\text{MMD}}(\theta, \Lambda) + \lambda_1 \mathcal{L}_{SupCon} + \lambda_2 \mathcal{L}_{TCon} + \lambda_3 \mathcal{L}_{Cross} + \lambda_{cls} \mathcal{L}_{cls}(\theta)$.
    $\theta' \leftarrow \theta - \eta \nabla_\theta \mathcal{L}_b(\theta, \phi)$.
  end for
  Extract θ': $z_i^t = E_t(x_i), \quad z_i^f = [a_i, p_i]$
  $\mathcal{L}_v(\theta', \psi) = -\mathbb{E}_{(z,d)}[\, d \log D_\psi(z) + (1-d)\log(1 - D_\psi(z))]$.
  $\psi \leftarrow \psi - \gamma \nabla_\psi \mathcal{L}_v(\theta', \psi)$.
  $g_v = \nabla_{\theta'} \mathcal{L}_v(\theta'); \quad \theta^+ = \theta' + \epsilon g_v, \quad \theta^- = \theta' - \epsilon g_v$.
  $\nabla_\phi \mathcal{L}_v \approx \frac{\eta}{2\epsilon}[\nabla_\phi \mathcal{L}_b(\theta^-, \phi) - \nabla_\phi \mathcal{L}_b(\theta^+, \phi)]$.
  $\phi \leftarrow \phi + \beta \nabla_\phi \mathcal{L}_v$.
end for

F DETAILS OF EXPERIMENT SETTING

F.1 DATASET DETAILS

We evaluate our method on eight datasets across four tasks: human activity recognition (HAR), gesture recognition (GR), sleep stage classification (SSC), and motor imagery (MI) task. The detailed information of the dataset is shown in Table 4.

- HHAR_SA Stisen et al. (2015): 30 participants, 3-axis accelerometer at 50 Hz, six activities including biking, sitting, standing, walking, walking upstairs, walking downstairs. We use non-overlapping 128-step segments as input.
- HHAR_D Gagnon-Audet et al. (2022): Unlike HHAR_SA, it is split into five device-model domains, corresponding to four smartphones and one smartwatch.

Table 4: Dataset statistics

Dataset	Length	Channels	Classes	Subjects	Train Size	Test Size	Task
HAR	128	9	6	30	2300	990	HAR
HHAR_SA	128	3	6	9	12716	5218	HAR
WISDM	128	3	6	36	1350	720	HAR
HHAR_D	500	6	6	5	10940	2734	HAR
EMG	200	8	6	4	5506	1377	GR
Sleep-EDF	3000	1	5	20	14280	6310	SSC
CAP	3000	19	6	5	32312	8078	SSC
PCL	500	48	2	3	18078	4520	MI

- HAR Garcia-Gonzalez et al. (2020): Tri-axial accelerometer, gyroscope, and body acceleration from 30 volunteers at 50 Hz, segmented into non-overlapping 128-step windows with six activities: walking, upstairs, downstairs, sitting, standing, and lying down.
- WISDM Kwapisz et al. (2011): It 3-axis accelerometer readings from 30 subjects, collected at 20 Hz. It contains 128-step non-overlapping windows to predict one of six activities: walking, jogging, sitting, standing, walking upstairs, and walking downstairs.
- EMG Lu et al. (2022): Surface electromyography (sEMG) from forearm muscles via an 8-sensor bracelet for gesture recognition, with labels comprising a 4-class set (left, right, up, down) and 2 static gestures.
- Sleep-EDF Goldberger et al. (2000): EEG from 20 individuals with 197 whole-night polysomnographic records as a sleep-staging benchmark. It uses the Fpz-Cz channel, with expert annotations in 30-second epochs over five stages: W, N1, N2, N3, and REM.
- CAP Goldberger et al. (2000): 108 polysomnographic recordings with 3 EEG channels, chin and bilateral anterior tibial EMG, airflow, respiratory effort, and ECG. Five device domains with six sleep stages: W, N1, N2, N3, REM, and one additional stage.
- PCL Cho et al. (2017): EEG for motor imagery (MI) classification from 3 subjects across 22 sessions and 14 channels, with three source domains from different research groups performing the same MI task and two classes: left-hand and right-hand MI.

Table 5: Comparison of Domain Adaptation Baselines in Four Aspects

Method	**Application**	**Category**	**Classifier Training**	**Distribution**
CDAN	UDA	Adversarial	Source-Only	Joint
HoMM	UDA	Statistical	Source-Only	Marginal
MMDA	UDA	Statistical	Source-Only	Marginal
DSAN	UDA	Statistical	Both	Joint
JMMD-HSIC	UDA	Statistical	Both	Joint
CoDATS	TSC-UDA	Adversarial	Source-Only	Marginal
RAINCOAT	TSC-UDA	Statistical	Source-Only	Marginal
AdvSKM	TSC-UDA	Statistical	Both	Marginal
CALDA	TSC-UDA	Adversarial & Contrastive	Source-Only	Marginal
CLUDA	TSC-UDA	Adversarial & Contrastive	Source-Only	Joint
CoTMix	TSC-UDA	Contrastive	Source-Only	Joint
ConFGD	TSC-UDA	Contrastive & Adversarial	Source-Only	Marginal
ACON	TSC-UDA	Adversarial	Source-Only	Marginal
GLA-DA	TSC-UDA	Adversarial	Both	Joint
EDEN	TSC-UDA	Adversarial	Both	Marginal
META (ours)	TSC-UDA	Contrastive & Statistical	Source-Only	Joint

F.2 Baseline Details

We select the following baseline methods: traditional UDA methods including CDAN Long et al. (2018), HoMM Chen et al. (2020), MMDA Rahman et al. (2020), DSAN Zhu et al. (2020), JMMD-HSIC Wang et al. (2026), as well as TSC-UDA methods specifically designed for time series, including CoDATS Wilson et al. (2020), RAINCOAT He et al. (2023), AdvSKM Liu & Xue (2021), CALDA Wilson et al. (2023), CLUDA Ozyurt et al. (2022), CoTMix Jia et al. (2025), ConFGD Guo et al. (2025), ACON Liu et al. (2024), GLA-DA Tu et al. (2024) and EDEN Liu et al. (2026). Table 5 summarizes these baseline methods across four aspects: Application, Category, Classifier Training, and Distribution Alignment.

- Application: Whether a method is designed for conventional UDA or specifically tailored for time series (TSC-UDA).
- Category: Whether the method adopts adversarial learning, statistical divergence, Self-supervised or emerging contrastive learning.
- Classifier Training: Whether the classifier is trained solely on source domain data or also leverages target domain data.
- Distribution: Whether the method aligns only the marginal distributions $p_s(x)$ or aligns the joint distribution with features and labels $p(x, y)$ or $p(y|x)$.

F.3 Implementation Details

In this section, we provide implementation details of META and the baseline methods. For each baseline method, we run each of the 10 cross-domain scenarios 5 times with different random seeds, reporting both the mean and standard deviation across runs for each scenario, and then report the average accuracy and its standard deviation across all 10 scenarios.

The implementation was done in PyTorch. The experiments were conducted on a NVIDIA GeForce RTX 4090 graphics card. We train feature extractors and classifiers for UDA methods and TSC-UDA methods. To ensure a fair comparison, all UDA methods adopt the same 1D-CNN encoder as feature extractors. For TSC-UDA methods, we adopt the feature extractors originally designed in their respective approaches. This backbone was selected because it consistently outperforms more complex alternatives such as 1D-ResNet-18 and TCN in benchmark evaluations, thereby attributing any performance differences solely to the adaptation algorithms. In addition, we used Adam for all methods, with learning rates individually tuned via grid search over $[10^{-4}, 10^{-1}]$ on source validation sets. For META, the main hyperparameters are listed in Tables 6.

Table 6: Hyperparameter Settings of META

Hyperparameter	HHAR_SA	HHAR_D	WISDM	HAR	EMG	Sleep-EDF	CAP	PCL
Epoch	60	50	60	60	80	80	30	40
Batch Size	32	32	32	32	32	32	32	32
Learning rate	1e-2	3e-3	5e-3	1e-3	1e-3	1e-2	1e-3	5e-4

G Full Results of META

In this section, we present comprehensive results of our closed-set domain adaptation experiments on eight public datasets spanning four distinct tasks: human activity recognition (HAR, including HHAR_SA, HHAR_D, WISDM, and HAR), sleep stage classification (SSC, including Sleep-EDF and CAP), gesture recognition (GR including EMG), and motion intent (MI including PCL). Table 7 summarizes the average Macro-F1 scores and the corresponding standard deviations over all domain pairs for each evaluated method. It is clearly observed that META consistently achieves the highest average Macro-F1 on every dataset compared to the second-best method in each case. To facilitate a thorough investigation, we provide the detailed per-pair accuracy (%) in Tables 8–15 and the corresponding Macro-F1 values in Tables 16–23.

Table 7: Mean and standard deviation of Macro-F1 score across eight datasets (HAR, SSC, GR, MI) over domain-pairs and 5 independent runs with random seeds. The **best** results are in bold, and the second-best results are underlined.

Task	HAR				SSC		GR	MI
Method	HHAR_SA	HHAR_D	WISDM	HAR	Sleep-EDF	CAP	EMG	PCL
Source	0.53 ± 0.19	0.42 ± 0.17	0.44 ± 0.14	0.79 ± 0.09	0.38 ± 0.13	0.52 ± 0.13	0.73 ± 0.06	0.59 ± 0.04
CDAN	0.73 ± 0.23	0.53 ± 0.16	0.51 ± 0.21	0.94 ± 0.07	0.64±0.08	0.55 ± 0.07	0.78 ± 0.05	0.60 ± 0.04
HoMM	0.65 ± 0.22	0.49 ± 0.17	0.46 ± 0.12	0.90 ± 0.09	0.64±0.10	0.58 ± 0.11	0.78 ± 0.04	0.63±0.05
MMDA	0.67 ± 0.26	0.48 ± 0.16	0.50 ± 0.14	0.88 ± 0.15	0.63 ± 0.12	0.54 ± 0.15	0.78 ± 0.04	0.41 ± 0.09
DSAN	0.71 ± 0.25	0.50 ± 0.15	0.53 ± 0.16	0.78 ± 0.19	0.55 ± 0.13	0.54 ± 0.10	0.78 ± 0.05	0.62 ± 0.04
JMMD-HSIC	0.62 ± 0.23	0.45 ± 0.17	0.52 ± 0.19	0.89 ± 0.11	0.63 ± 0.10	0.56 ± 0.10	0.77 ± 0.04	0.63±0.05
CoDATS	0.73 ± 0.24	0.52 ± 0.18	0.53 ± 0.20	0.92 ± 0.11	0.55 ± 0.12	0.61 ± 0.09	0.78 ± 0.06	0.63±0.05
RAINCOAT	0.74 ± 0.22	0.50 ± 0.20	0.70 ± 0.10	0.94 ± 0.07	0.63 ± 0.10	0.53 ± 0.08	0.78 ± 0.05	0.61 ± 0.05
AdvSKM	0.59 ± 0.22	0.45 ± 0.17	0.49 ± 0.12	0.87 ± 0.12	0.60 ± 0.12	0.55 ± 0.10	0.74 ± 0.06	**0.64±0.04**
CALDA	0.68 ± 0.19	0.53 ± 0.18	0.57 ± 0.22	0.92 ± 0.09	0.62 ± 0.10	0.57 ± 0.09	0.65 ± 0.05	0.51 ± 0.09
CLUDA	0.68 ± 0.20	0.52 ± 0.18	0.55 ± 0.19	0.93 ± 0.11	0.62 ± 0.09	0.51 ± 0.09	0.78 ± 0.04	0.58 ± 0.14
CoTMix	0.61 ± 0.15	0.44 ± 0.21	0.56 ± 0.21	0.80 ± 0.15	0.59 ± 0.10	0.46 ± 0.15	0.74 ± 0.05	0.43 ± 0.13
ConFGD	0.79 ± 0.21	0.43 ± 0.16	0.64 ± 0.20	0.96±0.04	0.31 ± 0.11	0.52 ± 0.14	0.77 ± 0.04	0.63±0.05
ACON	0.77 ± 0.21	0.60 ± 0.18	0.74±0.15	0.96±0.06	0.51 ± 0.11	0.64±0.05	0.83±0.06	0.63±0.04
GLA-DA	0.67 ± 0.22	0.44 ± 0.21	0.46 ± 0.12	0.89 ± 0.11	0.61 ± 0.08	0.56 ± 0.11	0.76 ± 0.04	0.62 ± 0.05
EDEN	0.81±0.18	0.62±0.15	0.74±0.25	**0.98±0.04**	0.52 ± 0.13	0.47 ± 0.10	0.82 ± 0.05	0.52 ± 0.10
META	**0.82±0.18**	**0.67±0.20**	**0.82±0.23**	**0.98±0.03**	**0.72±0.06**	**0.67±0.06**	**0.85±0.04**	**0.64±0.05**

Table 8: Accuracy (%) of each method on HHAR_SA across 10 domain transfer scenarios.

Method	0→2	1→6	2→4	4→0	4→5	5→1	5→2	7→2	7→5	8→4	Avg
Source	59.0 ± 5.2	81.7 ± 5.1	47.7 ± 2.4	30.8 ± 1.1	73.5 ± 3.6	77.7 ± 2.7	34.0 ± 2.1	36.2 ± 2.6	55.9 ± 2.8	72.7 ± 1.0	56.9
CDAN	**88.3±4.2**	93.8 ± 0.2	64.2 ± 4.1	46.7 ± 6.0	97.3 ± 1.2	97.6 ± 0.1	43.2 ± 5.9	40.8 ± 3.7	85.8 ± 5.5	76.3 ± 13.0	73.4
HoMM	73.7 ± 2.5	90.6 ± 0.6	54.9 ± 3.4	39.4 ± 1.1	88.9 ± 1.1	90.8 ± 0.8	39.8 ± 3.6	36.8 ± 0.3	66.4 ± 1.4	79.9 ± 5.7	66.1
MMDA	78.9 ± 0.5	92.8 ± 0.4	52.6 ± 6.2	32.6 ± 3.5	95.8 ± 0.4	93.4 ± 0.4	32.9 ± 1.9	40.6 ± 3.2	68.6 ± 6.1	89.5 ± 0.6	67.8
DSAN	72.2 ± 2.6	94.8±0.4	63.5 ± 1.1	37.5 ± 9.9	97.0 ± 0.4	**98.6±0.3**	39.0 ± 12.1	37.1 ± 1.1	84.9 ± 9.1	89.2 ± 12.0	71.4
JMMD-HSIC	73.0 ± 1.0	89.7 ± 1.2	54.9 ± 1.2	30.7 ± 3.2	87.0 ± 2.3	90.2 ± 0.7	36.8 ± 4.6	37.6 ± 1.1	64.1 ± 7.5	78.5 ± 5.1	64.3
CoDATS	71.3 ± 3.8	94.3 ± 0.4	60.8 ± 1.3	40.0 ± 3.9	96.6 ± 0.6	97.5 ± 0.3	52.0 ± 7.8	45.2 ± 1.3	80.5 ± 4.9	**97.6±0.7**	73.6
RAINCOAT	85.4 ± 1.5	93.6 ± 0.3	69.4 ± 14.4	46.5 ± 7.0	97.6 ± 0.5	97.3 ± 0.4	47.2 ± 3.3	43.4 ± 8.2	89.3 ± 1.6	75.4 ± 5.7	74.5
AdvSKM	72.3 ± 1.9	88.5 ± 0.9	50.4 ± 5.7	32.1 ± 3.2	83.1 ± 7.6	84.2 ± 4.2	36.3 ± 4.6	37.3 ± 1.6	57.0 ± 4.2	70.1 ± 7.7	61.1
CALDA	74.8 ± 1.9	91.5 ± 0.0	72.5 ± 1.3	49.7 ± 3.6	86.6 ± 2.6	94.1 ± 1.1	48.4 ± 5.4	38.7 ± 5.5	72.3 ± 1.1	77.6 ± 2.4	70.6
CLUDA	69.0 ± 1.9	91.0 ± 1.1	64.3 ± 0.0	52.1 ± 8.4	91.0 ± 5.0	93.2 ± 0.3	40.2 ± 0.0	51.7 ± 12.5	64.6 ± 6.8	74.4 ± 4.9	69.1
CoTMix	75.7 ± 3.5	83.1 ± 5.9	53.9 ± 2.8	51.0 ± 3.6	76.2 ± 4.4	84.4 ± 1.0	57.8 ± 1.6	54.3±8.0	65.3 ± 4.1	66.1 ± 5.3	66.8
ConFGD	85.9 ± 1.2	**94.9±0.0**	**83.3±0.1**	51.1 ± 3.2	97.1 ± 0.1	98.2±0.0	51.1 ± 4.3	58.9 ± 2.1	91.4±0.4	90.6 ± 0.0	80.3
ACON	86.6 ± 2.5	94.2 ± 8.1	77.7±7.9	52.2±0.9	**98.2±4.1**	98.0±3.0	59.6±1.6	46.7 ± 3.1	88.7 ± 2.4	90.0 ± 6.1	79.2
GLA-DA	71.3 ± 0.5	92.1 ± 0.0	62.3 ± 5.9	41.2 ± 2.1	91.2 ± 0.3	91.5 ± 1.0	42.4 ± 2.8	37.7 ± 0.6	69.4 ± 1.0	82.1 ± 4.1	68.1
EDEN	88.2±0.1	**94.9±1.2**	**83.3±3.1**	51.1 ± 0.1	97.1 ± 0.0	98.2±0.0	62.5±1.2	**69.2±2.1**	91.4±0.2	90.6 ± 0.0	82.6
META	88.2±2.5	94.6 ± 0.1	79.4±3.7	**54.7±3.5**	98.0±0.2	97.7 ± 0.1	**66.7±4.1**	60.3±0.3	**95.5±0.5**	97.1±0.3	**83.2**

Table 9: Accuracy (%) of each method on HHAR_D across 10 domain transfer scenarios.

Method	0→1	0→2	0→3	0→4	1→0	1→3	1→4	2→1	3→4	4→1	Avg
Source	67.9 ± 3.6	41.6 ± 5.0	38.7 ± 5.9	45.5 ± 6.2	33.4 ± 2.6	39.7 ± 3.4	40.5 ± 3.8	40.6 ± 3.6	83.2 ± 5.1	41.4 ± 2.4	47.3
CDAN	70.4 ± 2.1	54.7 ± 5.8	52.8±9.9	56.3±3.7	44.8 ± 7.0	53.4 ± 1.0	58.5 ± 4.6	42.2 ± 2.7	91.9±2.0	53.3 ± 4.0	57.8
HoMM	68.1 ± 4.2	46.6 ± 0.7	44.5 ± 0.4	52.3 ± 1.5	38.3 ± 1.1	60.1 ± 11.4	60.3 ± 6.0	47.0 ± 0.8	87.5 ± 0.6	49.5 ± 4.1	55.4
MMDA	67.2 ± 4.6	44.4 ± 1.7	40.5 ± 0.8	49.1 ± 1.6	37.5 ± 4.4	52.1 ± 2.0	60.9 ± 1.9	47.4±2.7	86.1 ± 1.3	46.8 ± 0.8	53.2
DSAN	71.6 ± 1.0	53.6 ± 0.2	45.2 ± 0.8	49.5 ± 5.2	49.2 ± 1.1	50.6 ± 0.5	49.5 ± 4.9	34.7 ± 1.9	84.8 ± 1.2	42.7 ± 0.5	53.1
JMMD-HSIC	66.1 ± 1.4	44.6 ± 3.9	43.8 ± 4.0	52.5 ± 6.2	38.5 ± 7.4	48.4 ± 8.5	53.1 ± 9.4	39.6 ± 5.0	84.8 ± 1.5	46.6 ± 2.4	51.8
CoDATS	69.4 ± 3.2	50.6 ± 4.7	46.5 ± 16.0	42.1 ± 12.4	57.8±0.0	67.2 ± 4.2	64.2 ± 12.4	38.8 ± 2.2	83.5 ± 1.8	35.6 ± 2.0	55.6
RAINCOAT	69.5 ± 3.1	35.8 ± 7.2	32.6 ± 4.0	42.3 ± 6.2	62.5±2.2	53.5 ± 4.5	60.0 ± 1.2	42.8 ± 12.6	87.5 ± 1.8	51.3 ± 0.8	53.8
AdvSKM	66.9 ± 1.9	43.9 ± 4.7	40.6 ± 4.2	48.9 ± 1.0	35.2 ± 1.1	47.7 ± 0.4	46.7 ± 0.9	44.8 ± 0.2	87.1 ± 0.3	46.5 ± 3.2	50.8
CALDA	64.9 ± 1.0	40.9 ± 2.9	43.0 ± 10.4	45.0 ± 3.5	53.1 ± 2.2	77.8 ± 1.0	54.7 ± 4.9	45.4 ± 4.1	86.4 ± 2.8	46.2 ± 1.0	55.7
CLUDA	69.7 ± 0.7	41.5 ± 5.4	33.7 ± 0.4	50.4 ± 10.9	56.3 ± 0.0	58.5 ± 7.6	52.4 ± 2.8	39.5 ± 3.2	86.4 ± 3.1	50.0 ± 11.2	53.8
CoTMix	70.0 ± 0.0	34.5 ± 2.8	31.1 ± 0.9	35.7 ± 5.2	46.9 ± 8.8	34.9 ± 2.9	47.1 ± 10.0	47.8±2.4	89.6 ± 0.6	54.3 ± 1.7	49.2
ConFGD	61.2 ± 0.2	47.8 ± 2.1	47.5 ± 1.0	54.9 ± 2.3	31.3 ± 1.5	37.7 ± 4.5	40.8 ± 1.0	35.7 ± 2.1	83.7 ± 0.1	44.8 ± 4.1	48.5
ACON	69.0 ± 1.2	57.9±0.0	44.6 ± 0.9	49.2 ± 7.1	62.5±0.5	80.5±3.2	84.6±2.5	45.0 ± 0.4	88.1 ± 2.1	55.3±0.8	63.7
GLA-DA	51.2 ± 0.6	31.3 ± 4.8	31.6 ± 1.2	33.0 ± 8.5	55.5 ± 7.7	46.9 ± 1.1	46.9 ± 0.9	40.6 ± 1.1	87.6 ± 0.2	51.0 ± 14.0	47.6
EDEN	79.8±0.0	61.7±1.2	63.8±2.1	64.6±0.2	73.4±1.0	64.8 ± 1.6	68.3 ± 2.5	**57.5±1.0**	90.2 ± 0.6	**63.5±4.1**	68.8
META	**80.3±1.3**	**65.0±3.1**	**66.8±2.4**	**66.0±5.2**	**75.0±0.2**	**84.9±1.2**	**86.7±0.0**	48.1±2.9	**97.1±0.1**	51.0 ± 1.2	**72.1**

Table 10: Accuracy (%) of each method on WISDM across 10 domain transfer scenarios.

Method	2→32	4→15	7→30	12→7	12→19	18→20	20→30	21→31	25→29	26→2	Avg
Source	81.2 ± 6.3	79.2 ± 9.5	63.4 ± 5.6	63.2 ± 12.9	55.8 ± 5.2	64.2 ± 9.2	58.3 ± 10.1	35.7 ± 12.5	62.2 ± 1.0	60.6 ± 3.7	62.4
CDAN	73.2 ± 21.5	70.8 ± 5.9	86.4 ± 1.4	76.6 ± 0.7	50.8 ± 3.2	69.5 ± 1.7	82.0 ± 8.9	56.6 ± 5.1	50.8 ± 1.2	82.3 ± 4.3	69.9
HoMM	74.4 ± 5.9	72.9 ± 3.6	78.6 ± 2.6	63.2 ± 3.3	51.3 ± 5.8	75.2 ± 5.8	77.7 ± 1.0	55.4 ± 1.2	62.2 ± 1.9	70.3 ± 4.6	68.1
MMDA	72.5 ± 4.1	76.0 ± 1.5	86.4 ± 4.1	70.3 ± 9.6	79.2 ± 2.7	75.0 ± 11.2	81.1 ± 3.4	66.9 ± 6.0	61.7 ± 7.1	80.5 ± 1.7	75.0
DSAN	73.4 ± 7.1	77.8 ± 7.3	81.2 ± 2.4	70.1 ± 7.4	60.1 ± 1.6	84.6 ± 0.7	77.3 ± 2.0	63.9 ± 8.3	60.0 ± 2.9	81.3 ± 9.2	73.0
JMMD-HSIC	77.3 ± 2.2	67.4 ± 12.0	87.7 ± 0.6	75.0 ± 2.8	51.3 ± 6.8	71.5 ± 9.5	82.2 ± 3.1	58.6 ± 1.8	57.8 ± 1.9	76.0 ± 1.9	70.5
CoDATS	78.3 ± 1.4	73.6 ± 12.6	89.6 ± 3.0	69.4 ± 3.0	57.3 ± 1.9	86.6 ± 1.2	77.3 ± 4.9	55.0 ± 0.7	61.7 ± 6.0	87.0 ± 3.5	73.6
RAINCOAT	79.7 ± 2.9	81.9 ± 8.7	86.7 ± 2.2	75.7 ± 5.3	63.4 ± 9.9	86.2 ± 2.5	83.2 ± 2.4	61.4 ± 6.7	54.4 ± 2.5	88.6 ± 3.1	76.1
AdvSKM	76.3 ± 3.0	81.9 ± 9.8	78.6 ± 3.5	65.6 ± 13.7	64.1 ± 2.2	78.0 ± 4.2	72.5 ± 7.5	55.8 ± 1.8	60.6 ± 3.5	66.7 ± 3.9	70.0
CALDA	80.3 ± 9.3	77.1 ± 9.9	90.5±3.9	80.6 ± 6.4	70.2 ± 10.4	84.6 ± 3.3	81.9 ± 2.3	64.6 ± 7.9	25.0 ± 11.8	87.1 ± 2.5	74.2
CLUDA	77.3 ± 0.8	79.9 ± 4.8	86.7 ± 1.5	79.9 ± 4.7	54.0 ± 7.0	86.2 ± 3.5	80.9 ± 3.0	57.0 ± 1.8	62.2 ± 2.5	73.2 ± 3.2	73.7
CoTMix	82.1 ± 3.0	91.0±1.2	90.3 ± 3.4	81.9 ± 5.3	60.4 ± 2.3	62.6 ± 1.9	76.7 ± 5.9	47.4 ± 2.8	55.0 ± 0.0	86.2 ± 6.9	73.4
ConFGD	84.1 ± 1.2	91.7 ± 0.9	92.2 ± 0.1	80.2 ± 2.0	84.8±3.1	87.8 ± 0.2	74.8 ± 4.6	57.8 ± 3.0	65.0±1.0	89.0 ± 0.0	80.7
ACON	92.8±1.5	81.3 ± 2.5	96.1±5.7	84.4±7.4	75.8±2.2	90.2±2.5	91.3±10.4	71.1±5.3	63.3±3.2	95.1±2.8	84.1
GLA-DA	64.5 ± 1.0	79.2 ± 11.8	72.8 ± 6.9	66.1 ± 14.0	58.7 ± 7.0	71.3 ± 2.6	62.6 ± 4.8	59.6 ± 14.5	59.2 ± 3.5	72.6 ± 2.6	66.7
EDEN	78.3 ± 1.0	**97.9±0.2**	99.0±0.7	89.6±0.0	75.8 ± 2.1	86.6 ± 4.3	88.3 ± 1.4	**80.7±8.5**	56.7 ± 2.4	**100.0±2.6**	85.3
META	**94.2±1.0**	95.8±0.0	**100.0±0.0**	**96.9±0.0**	**88.6±2.1**	**91.5±0.9**	**95.1±0.8**	78.3±4.1	**66.7±0.0**	**100.0±0.0**	**90.7**

Table 11: Accuracy (%) of each method on HAR across 10 domain transfer scenarios.

Method	2→11	6→23	7→13	9→18	12→16	13→19	18→21	20→6	23→13	24→12	Avg
Source	81.3 ± 4.7	75.5 ± 4.5	84.4 ± 5.4	75.0 ± 6.3	71.4 ± 4.8	81.8 ± 4.8	99.0 ± 0.9	85.9 ± 3.1	72.9 ± 7.0	89.1 ± 6.3	81.6
CDAN	**100.0±0.0**	95.3 ± 0.7	91.1 ± 0.7	81.8 ± 3.7	83.3±4.4	**100.0±0.0**	**100.0±0.0**	**100.0±0.0**	93.2±3.7	99.5 ± 0.7	94.4
HoMM	**100.0±0.0**	87.5 ± 4.5	90.3 ± 2.6	78.8 ± 1.6	78.5 ± 6.9	98.6 ± 0.6	99.3 ± 0.6	89.2 ± 7.7	86.8 ± 2.2	97.6 ± 2.6	90.7
MMDA	**100.0±0.0**	95.5 ± 0.6	87.8 ± 5.9	66.7 ± 6.8	68.1 ± 0.6	**100.0±0.0**	**100.0±0.0**	99.7±0.6	76.4 ± 5.7	98.3 ± 1.6	89.2
DSAN	92.2 ± 10.8	72.2 ± 15.2	73.3 ± 16.2	68.1 ± 10.3	55.2 ± 3.8	98.6 ± 1.6	**100.0±0.0**	87.5 ± 10.9	71.5 ± 3.0	96.9 ± 2.1	81.6
JMMD-HSIC	98.4±1.6	87.5 ± 1.8	92.0 ± 0.6	79.2 ± 1.8	70.8 ± 2.8	99.0±0.0	**100.0±0.0**	89.9 ± 6.3	80.9 ± 11.0	98.3 ± 1.2	89.6
CoDATS	**100.0±0.0**	95.1 ± 0.6	92.7 ± 1.0	79.9 ± 2.6	69.8 ± 0.0	**100.0±0.0**	**100.0±0.0**	99.0 ± 1.0	87.2 ± 3.0	99.7±0.6	92.3
RAINCOAT	**100.0±0.0**	95.5 ± 0.6	92.0 ± 1.6	82.6 ± 3.0	80.9 ± 5.2	**100.0±0.0**	**100.0±0.0**	99.0 ± 1.8	91.7 ± 3.8	99.7±0.6	94.1
AdvSKM	98.4±2.7	86.1 ± 3.9	91.7 ± 0.0	76.0 ± 5.4	69.8 ± 2.8	94.8 ± 9.0	99.7±0.6	84.7 ± 7.7	83.3 ± 8.1	99.7±0.6	88.4
CALDA	96.9 ± 4.4	90.6 ± 2.9	93.2 ± 0.7	71.9 ± 10.3	78.6 ± 2.2	**100.0±0.0**	**100.0±0.0**	98.4 ± 0.7	92.2 ± 3.7	98.4 ± 0.7	92.0
CLUDA	**100.0±0.0**	95.8 ± 0.0	93.8±0.0	81.3 ± 4.2	71.9 ± 2.8	**100.0±0.0**	**100.0±0.0**	**100.0±0.0**	90.3 ± 6.0	**100.0±0.0**	93.3
CoTMix	76.0 ± 9.9	57.6 ± 6.4	89.2 ± 0.6	70.1 ± 5.7	82.3 ± 2.8	80.6 ± 5.7	98.6 ± 1.2	92.4 ± 3.3	85.8 ± 4.2	88.9 ± 0.6	82.2
ConFGD	**100.0±0.0**	90.6 ± 7.4	95.3 ± 2.2	**96.4±2.2**	88.0±0.7	**100.0±0.0**	**100.0±0.0**	99.5 ± 0.7	91.1 ± 0.7	**100.0±0.0**	96.1
ACON	**100.0±1.0**	96.9±4.3	96.9±6.1	88.5±4.1	81.3 ± 2.1	**100.0±0.0**	**100.0±0.0**	**100.0±0.2**	91.7 ± 1.5	**100.0±0.0**	95.5
GLA-DA	97.9 ± 1.8	87.5 ± 1.8	92.0 ± 0.6	79.2 ± 1.8	71.5 ± 2.2	99.0±0.0	**100.0±0.0**	89.9 ± 6.3	80.2 ± 9.9	98.3 ± 1.2	89.5
EDEN	**100.0±0.0**	97.9±1.5	99.0±0.7	93.8 ± 1.5	84.4 ± 0.0	**100.0±0.0**	**100.0±0.0**	**100.0±0.0**	99.0±1.7	**100.0±0.0**	97.4
META	**100.0±0.0**	**100.0±0.0**	**100.0±0.0**	95.8±0.9	**91.7±2.4**	**100.0±0.0**	**100.0±0.0**	**100.0±0.0**	**100.0±0.0**	**100.0±0.0**	**98.8**

Table 12: Accuracy (%) of each method on Sleep-EDF across 10 domain transfer scenarios.

Method	9→11	2→5	12→5	7→18	16→1	9→14	4→12	10→7	6→3	8→10	Avg
Source	65.1 ± 4.8	60.4 ± 4.0	48.7 ± 1.8	71.5 ± 1.9	66.3 ± 2.2	55.7 ± 7.7	47.9 ± 11.9	50.5 ± 0.6	60.1 ± 3.2	43.6 ± 12.2	57.0
CDAN	66.6 ± 4.1	72.8 ± 0.1	80.8 ± 2.1	73.4 ± 2.5	78.8 ± 2.8	84.1 ± 2.1	60.6 ± 3.7	79.9±0.3	84.8 ± 0.6	71.8±1.2	75.4
HoMM	71.5 ± 2.8	71.3 ± 1.6	83.2±2.7	68.6 ± 8.6	75.2 ± 2.0	84.5 ± 3.7	55.8 ± 4.5	79.8 ± 2.1	84.2 ± 0.6	68.0 ± 2.3	74.2
MMDA	63.4 ± 2.6	72.7 ± 0.5	81.8 ± 0.7	75.2 ± 0.3	77.0 ± 1.2	83.7 ± 3.2	61.4 ± 1.8	80.6±1.5	85.3 ± 0.7	66.1 ± 4.9	74.7
DSAN	67.9 ± 4.8	66.8 ± 0.0	82.5 ± 0.7	76.6 ± 1.7	72.9 ± 3.0	83.9 ± 4.6	68.5 ± 0.1	67.5 ± 1.3	79.5 ± 2.8	57.4 ± 1.7	72.3
JMMD-HSIC	69.6 ± 1.3	70.4 ± 1.9	81.8 ± 1.9	75.4 ± 0.5	76.5 ± 1.6	84.6 ± 2.7	62.0 ± 0.2	76.5 ± 2.1	84.9 ± 1.5	65.0 ± 5.6	74.7
CoDATS	63.9 ± 2.5	69.8 ± 4.0	78.1 ± 3.3	79.8±0.7	78.0 ± 5.5	78.1 ± 13.8	73.2±3.9	62.8 ± 2.9	64.7 ± 9.8	62.1 ± 4.5	71.1
RAINCOAT	68.2 ± 0.3	74.1 ± 4.3	82.0 ± 0.3	78.3 ± 0.6	77.3 ± 0.8	87.0±1.0	71.6 ± 1.2	74.5 ± 0.6	80.0 ± 2.1	59.2 ± 1.8	75.2
AdvSKM	59.5 ± 2.3	69.0 ± 1.0	77.6 ± 1.0	60.0 ± 0.0	72.4 ± 3.0	80.4 ± 3.5	65.5 ± 1.8	77.8 ± 1.1	81.3 ± 2.3	49.0 ± 0.6	69.3
CALDA	75.4±2.2	68.3 ± 1.8	81.5 ± 1.5	78.2 ± 0.7	78.2 ± 0.6	84.0 ± 1.9	69.3 ± 0.8	77.3 ± 2.3	84.6 ± 1.9	50.5 ± 5.9	74.7
CLUDA	64.6 ± 6.1	75.7±0.8	81.4 ± 1.8	73.7 ± 2.2	76.6 ± 3.9	84.0 ± 2.3	63.3 ± 7.2	72.7 ± 3.4	86.3±1.8	53.0 ± 5.7	73.1
CoTMix	59.8 ± 1.4	70.4 ± 0.4	72.4 ± 0.1	78.3 ± 0.3	76.4 ± 0.2	81.3 ± 1.2	58.9 ± 0.4	66.3 ± 3.2	80.5 ± 1.4	59.9 ± 3.2	70.4
ConFGD	57.8 ± 3.9	55.2 ± 3.7	47.2 ± 2.9	48.5 ± 2.1	43.7 ± 23.1	67.1 ± 4.6	47.1 ± 3.9	44.7 ± 3.8	72.9 ± 1.8	41.4 ± 28.1	52.6
ACON	48.8 ± 0.5	67.0 ± 1.3	78.1 ± 2.4	79.1 ± 0.5	80.3±0.0	65.2 ± 0.2	50.8 ± 0.2	64.1 ± 2.1	71.1 ± 1.4	64.5 ± 0.2	66.9
GLA-DA	68.5 ± 4.0	69.3 ± 5.0	80.1 ± 0.3	68.9 ± 3.0	71.6 ± 0.1	85.4 ± 5.4	63.0 ± 1.8	74.8 ± 2.4	85.0 ± 1.1	62.0 ± 5.0	72.9
EDEN	67.5 ± 3.7	63.6 ± 0.6	71.8 ± 4.3	76.5 ± 0.7	77.3 ± 3.0	79.4 ± 3.3	60.5 ± 2.8	55.6 ± 9.5	78.0 ± 0.1	56.0 ± 5.6	68.6
META	**83.2±2.0**	**82.3±3.7**	**85.1±3.0**	**84.6±2.8**	**85.7±2.7**	**89.4±1.6**	**77.3±2.2**	**84.7±3.4**	**87.3±0.3**	**76.4±2.8**	**83.6**

Table 13: Accuracy (%) of each method on CAP across 10 domain transfer scenarios.

Method	0→1	0→3	0→4	1→0	1→4	2→3	3→0	3→1	4→1	4→3	Avg
Source	55.9 ± 0.2	74.6 ± 0.7	71.7 ± 1.2	54.2 ± 1.4	62.2 ± 1.4	53.9 ± 8.1	71.4 ± 2.3	51.5 ± 3.9	50.9 ± 0.9	70.6 ± 1.4	61.7
CDAN	67.6 ± 0.3	75.5 ± 1.9	65.3 ± 3.4	68.6±4.8	72.4 ± 1.0	50.9 ± 9.0	74.0 ± 1.0	65.6 ± 7.1	66.1 ± 2.8	74.0 ± 2.0	68.0
HoMM	65.8 ± 1.1	77.0 ± 3.2	73.6 ± 3.0	63.5 ± 1.1	69.1 ± 4.1	56.9 ± 1.4	72.0 ± 2.9	55.7 ± 1.0	53.6 ± 0.7	72.3 ± 1.2	65.9
MMDA	53.3 ± 0.1	77.4 ± 1.2	71.9 ± 1.5	61.6 ± 2.1	71.0 ± 1.0	58.3 ± 0.9	70.7 ± 2.1	47.3 ± 2.1	54.2 ± 0.2	69.8 ± 0.5	63.6
DSAN	68.4 ± 3.1	71.1 ± 0.1	72.6 ± 0.8	64.2 ± 2.1	69.9 ± 3.7	48.6 ± 2.6	72.4 ± 1.2	64.4 ± 0.8	56.7 ± 4.1	74.3 ± 2.2	66.3
JMMD-HSIC	64.8 ± 2.1	76.7 ± 0.6	73.7 ± 2.3	58.6 ± 2.7	64.4 ± 1.8	57.1 ± 5.0	71.6 ± 0.9	54.0 ± 1.7	53.0 ± 2.9	72.7 ± 2.2	64.7
CoDATS	69.5 ± 1.0	77.8±0.5	76.9±0.7	66.7 ± 0.1	72.8 ± 2.4	55.9 ± 1.1	76.8 ± 1.2	64.8 ± 4.3	55.9 ± 1.3	75.6 ± 0.5	69.3
RAINCOAT	64.1 ± 2.0	70.5 ± 0.0	67.5 ± 1.6	62.7 ± 1.2	66.6 ± 0.9	38.1 ± 3.7	63.9 ± 2.0	64.6 ± 1.2	69.9±0.2	65.0 ± 1.3	63.3
AdvSKM	60.5 ± 1.5	76.1 ± 2.0	75.4 ± 3.9	59.5 ± 4.8	62.8 ± 2.1	56.1 ± 3.9	71.9 ± 0.3	56.6 ± 0.9	54.4 ± 1.3	71.4 ± 2.5	64.5
CALDA	65.7 ± 1.7	74.1 ± 3.0	75.0 ± 5.9	64.1 ± 4.3	68.9 ± 3.3	58.2 ± 2.9	77.0±4.8	70.5 ± 3.2	66.4 ± 2.7	75.2 ± 2.1	69.5
CLUDA	61.5 ± 2.1	69.1 ± 4.1	67.3 ± 0.4	67.0 ± 3.2	70.6 ± 2.2	56.4 ± 0.9	72.0 ± 0.4	53.8 ± 1.5	56.9 ± 1.2	68.9 ± 4.1	64.3
CoTMix	60.6 ± 0.3	74.5 ± 1.2	72.1 ± 3.2	46.8 ± 2.1	48.3 ± 1.3	42.9 ± 3.2	61.9 ± 0.5	53.1 ± 0.8	56.9 ± 2.3	68.2 ± 1.8	58.5
ConFGD	54.9 ± 0.1	76.4 ± 1.2	74.8 ± 1.0	56.6 ± 3.4	64.9 ± 4.1	**61.6±1.5**	70.9 ± 0.2	38.5 ± 4.7	48.0 ± 3.2	73.0 ± 0.9	62.0
ACON	74.1±0.0	76.9 ± 0.3	76.6 ± 1.3	68.2 ± 0.6	73.1±0.1	59.1 ± 1.0	75.3 ± 1.6	73.0±0.6	68.2 ± 0.1	76.3±1.2	72.1
GLA-DA	61.8 ± 0.2	76.4 ± 0.5	73.8 ± 1.2	56.6 ± 1.5	63.9 ± 4.8	57.6 ± 0.0	71.5 ± 1.1	51.0 ± 6.0	46.4 ± 12.1	72.4 ± 0.0	63.1
EDEN	39.6 ± 2.1	72.8 ± 1.2	73.8 ± 0.3	63.1 ± 1.5	64.8 ± 0.3	48.3 ± 2.6	72.0 ± 0.8	42.4 ± 1.3	44.6 ± 0.5	69.9 ± 1.2	59.1
META	**77.4±1.1**	**80.2±0.4**	**80.9±0.1**	**71.2±1.0**	**75.6±1.8**	61.3±2.5	**78.2±0.8**	**77.0±0.1**	**75.4±4.0**	**78.8±0.8**	**75.6**

Table 14: Accuracy (%) of each method on EMG across 10 domain transfer scenarios.

Method	0→1	0→2	0→3	1→2	1→3	2→0	2→1	2→3	3→1	3→2	Avg
Source	85.2 ± 3.2	71.1 ± 2.9	72.0 ± 0.7	68.1 ± 0.9	71.9 ± 0.9	64.7 ± 1.3	79.3 ± 1.2	70.8 ± 0.2	82.1 ± 0.0	71.7 ± 1.1	73.7
CDAN	85.5 ± 2.0	73.8 ± 0.4	77.3 ± 0.2	78.4 ± 0.9	78.9 ± 0.2	69.8 ± 3.3	83.1 ± 1.4	79.1 ± 2.7	81.3 ± 0.0	72.7 ± 1.5	78.0
HoMM	82.8 ± 2.2	73.9 ± 0.2	79.5 ± 1.5	77.7 ± 0.7	78.9 ± 0.2	71.4 ± 2.4	83.5 ± 2.0	76.1 ± 2.9	80.8 ± 1.8	74.4 ± 0.4	77.9
MMDA	83.5 ± 2.8	73.1 ± 1.3	78.9 ± 0.7	76.7 ± 0.2	78.4 ± 1.3	73.4±4.4	85.4 ± 2.6	75.9 ± 2.2	81.4 ± 0.2	74.7 ± 1.8	78.1
DSAN	86.4 ± 0.0	75.3 ± 2.2	78.4 ± 0.9	76.1 ± 0.2	76.6 ± 1.3	70.3 ± 2.2	83.2 ± 0.8	79.2 ± 2.9	82.2 ± 0.2	74.2 ± 1.5	78.2
JMMD-HSIC	82.6 ± 3.0	72.1 ± 0.4	76.6 ± 1.4	77.6 ± 2.1	78.0 ± 1.0	72.3 ± 1.4	82.4 ± 1.5	74.2 ± 2.1	81.1 ± 1.6	73.5 ± 3.1	77.0
CoDATS	86.6 ± 1.2	74.5 ± 0.7	75.9 ± 2.7	78.4 ± 1.8	80.5 ± 0.7	65.9 ± 3.5	84.2 ± 3.0	75.8 ± 1.5	79.5 ± 2.8	76.1 ± 0.2	77.8
RAINCOAT	84.4 ± 1.2	74.1 ± 0.0	75.9 ± 0.9	77.8 ± 0.4	80.5 ± 0.7	67.7 ± 0.2	84.5 ± 0.2	76.7 ± 0.7	82.7 ± 0.4	75.9 ± 0.9	78.0
AdvSKM	80.4 ± 0.8	71.9 ± 0.0	73.9 ± 1.1	77.7 ± 0.7	80.8 ± 1.5	63.3 ± 0.7	77.7 ± 4.2	70.8 ± 6.0	79.7 ± 1.4	70.9 ± 1.8	74.7
CALDA	69.0 ± 3.2	69.5 ± 1.5	62.2 ± 0.4	67.7 ± 0.2	65.3 ± 0.9	57.8 ± 4.4	63.8 ± 2.2	63.9 ± 0.2	74.9 ± 3.0	65.0 ± 3.1	65.9
CLUDA	84.1 ± 1.6	74.8 ± 0.7	72.7 ± 2.4	78.0 ± 0.2	78.0 ± 0.2	71.3 ± 0.0	83.1 ± 2.2	76.4 ± 0.2	81.7 ± 0.6	75.3 ± 1.8	77.5
CoTMix	79.1 ± 3.8	74.5 ± 1.1	66.6 ± 2.2	76.6 ± 0.0	79.1 ± 1.3	69.7 ± 0.4	79.0 ± 1.2	72.5 ± 3.1	77.4 ± 2.6	75.3 ± 1.8	75.0
ConFGD	84.7 ± 1.2	72.7 ± 1.1	73.8 ± 3.1	76.4 ± 0.7	80.2 ± 2.0	71.7 ± 2.9	82.4 ± 2.0	76.3 ± 2.2	80.4 ± 0.4	74.2 ± 2.9	77.3
ACON	**92.3±0.1**	80.3±0.0	81.6±1.2	81.6±0.3	84.4±0.9	70.9 ± 0.5	90.6±2.2	**82.2±1.5**	84.9±0.4	**80.3±0.0**	82.9
GLA-DA	81.5 ± 2.0	72.0 ± 0.2	73.3 ± 1.1	77.0 ± 1.1	78.0 ± 1.1	70.5 ± 2.9	80.8 ± 0.2	72.5 ± 1.8	81.7 ± 1.0	73.3 ± 0.7	76.1
EDEN	88.6 ± 0.3	78.1 ± 1.3	78.8 ± 2.1	82.2±0.3	84.4±1.2	78.8±0.9	90.9±0.2	80.0 ± 0.0	84.4 ± 0.5	75.6 ± 1.3	82.2
META	91.5±0.2	**80.9±1.3**	**83.8±0.7**	**82.8±0.1**	**85.9±2.9**	**80.3±4.6**	**91.5±0.5**	80.3±1.1	**87.2±0.0**	79.7±1.2	**84.4**

Table 15: Accuracy (%) of each method on PCL across 10 domain transfer scenarios.

Method	0→1	0→2	1→0	1→2	2→0	2→1	Avg
Source	60.6 ± 0.1	58.2 ± 3.1	59.2 ± 1.5	60.1 ± 0.5	56.5 ± 0.3	67.1 ± 0.1	60.3
CDAN	67.5 ± 0.4	59.0 ± 0.5	62.7 ± 0.2	62.0 ± 0.3	57.5 ± 1.1	61.3 ± 4.5	61.7
HoMM	68.8±0.1	57.9 ± 2.1	61.4 ± 4.6	64.1±1.2	59.2 ± 0.0	68.1 ± 0.2	63.2
MMDA	50.5 ± 1.0	49.6 ± 1.0	57.4 ± 0.0	49.8 ± 0.3	51.2 ± 0.2	55.1 ± 1.0	52.3
DSAN	68.1 ± 2.1	57.0 ± 1.0	63.0 ± 1.3	60.7 ± 1.2	58.2 ± 0.7	65.8 ± 1.3	62.1
JMMD-HSIC	67.7 ± 0.6	**61.1±0.3**	62.6 ± 0.3	63.4 ± 0.6	57.9 ± 0.2	68.9 ± 0.4	63.6
CoDATS	68.3 ± 0.4	59.7 ± 0.1	62.0 ± 1.0	61.8 ± 1.2	59.4 ± 0.5	**69.9±0.9**	63.5
RAINCOAT	60.7 ± 0.0	55.5 ± 0.6	62.4 ± 1.0	62.7 ± 0.8	58.7 ± 1.5	68.5 ± 1.0	61.4
AdvSKM	68.9±0.1	59.7 ± 0.9	63.1±1.9	62.9±0.3	59.5±1.4	68.9 ± 0.0	63.8
CALDA	61.8 ± 0.6	57.6 ± 0.7	62.9 ± 0.0	58.9 ± 1.0	50.4 ± 0.5	50.8 ± 0.7	57.1
CLUDA	**69.7±0.0**	51.4 ± 0.9	62.8 ± 1.2	63.6±1.4	54.9 ± 0.4	69.2 ± 0.1	61.9
CoTMix	51.0 ± 0.0	51.3 ± 1.0	64.7 ± 0.6	51.5 ± 0.3	49.8 ± 1.7	50.1 ± 2.1	53.1
ConFGD	68.2 ± 0.0	59.3 ± 0.1	61.9 ± 1.0	62.7 ± 0.0	57.8 ± 0.5	68.7 ± 0.1	63.1
ACON	68.2 ± 1.0	59.2 ± 2.1	61.7 ± 0.9	62.6 ± 1.3	**59.7±1.0**	67.4 ± 0.5	63.1
GLA-DA	67.2 ± 2.1	60.7 ± 0.3	61.5 ± 0.9	62.2 ± 0.2	57.6 ± 0.2	69.3±1.1	63.1
EDEN	55.9 ± 1.0	55.3 ± 0.1	61.3 ± 0.5	63.0 ± 1.2	56.1 ± 0.4	50.1 ± 0.1	56.9
META	69.3±1.5	61.0±1.6	**65.1±1.0**	**66.1±3.0**	59.1 ± 0.1	68.7 ± 0.5	**64.9**

Table 16: Macro-F1 Score of each method on HHAR_SA across 10 domain transfer scenarios.

Method	$0 \to 2$	$1 \to 6$	$2 \to 4$	$4 \to 0$	$4 \to 5$	$5 \to 1$	$5 \to 2$	$7 \to 2$	$7 \to 5$	$8 \to 4$	Avg
Source	0.53	0.78	0.34	0.32	0.71	0.77	0.28	0.38	0.52	0.65	0.53
CDAN	**0.87**	0.94	0.61	0.46	0.97	0.98	0.42	0.43	0.86	0.77	0.73
HoMM	0.69	0.91	0.47	0.39	0.89	0.91	0.39	0.39	0.66	0.79	0.65
MMDA	0.73	0.93	0.48	0.35	0.96	0.93	0.31	0.42	0.67	0.90	0.67
DSAN	0.67	**0.95**	0.64	0.37	0.97	**0.99**	0.38	0.40	0.85	0.89	0.71
JMMD-HSIC	0.68	0.90	0.45	0.31	0.87	0.90	0.33	0.40	0.63	0.77	0.62
CoDATS	0.67	0.94	0.61	0.39	0.97	0.98	0.49	0.45	0.80	**0.98**	0.73
RAINCOAT	0.84	0.94	0.68	0.45	**0.98**	0.97	0.45	0.45	0.90	0.75	0.74
AdvSKM	0.68	0.88	0.40	0.32	0.83	0.84	0.32	0.40	0.55	0.68	0.59
CALDA	0.67	0.92	0.65	0.45	0.86	0.94	0.46	0.41	0.73	0.76	0.68
CLUDA	0.66	0.91	0.58	0.52	0.91	0.93	0.36	0.50	0.63	0.74	0.68
CoTMix	0.73	0.80	0.41	0.46	0.72	0.84	0.50	0.49	0.59	0.56	0.61
ConFGD	0.85	**0.95**	**0.81**	0.48	0.97	0.98	0.46	0.56	0.92	0.89	0.79
ACON	0.86	0.94	0.70	0.49	**0.98**	0.98	0.55	0.45	0.89	0.90	0.77
GLA-DA	0.65	0.92	0.56	0.41	0.92	0.91	0.40	0.38	0.69	0.82	0.67
EDEN	**0.87**	**0.95**	**0.81**	0.48	0.97	0.98	0.57	**0.68**	0.92	0.89	0.81
META	**0.87**	**0.95**	0.71	**0.54**	**0.98**	0.98	**0.66**	0.59	**0.96**	0.97	**0.82**

Table 17: Macro-F1 Score of each method on HHAR_D across 10 domain transfer scenarios.

Method	$0 \to 1$	$0 \to 2$	$0 \to 3$	$0 \to 4$	$1 \to 0$	$1 \to 3$	$1 \to 4$	$2 \to 1$	$3 \to 4$	$4 \to 1$	Avg
Source	0.63	0.31	0.30	0.36	0.45	0.34	0.35	0.32	0.82	0.33	0.42
CDAN	0.67	0.45	0.41	0.45	0.55	0.48	0.55	0.36	0.92	0.49	0.53
HoMM	0.63	0.33	0.31	0.39	0.50	0.56	0.54	0.39	0.86	0.43	0.49
MMDA	0.62	0.32	0.30	0.37	0.48	0.48	0.56	0.42	0.85	0.43	0.48
DSAN	0.70	0.46	0.38	0.41	0.52	0.48	0.45	0.34	0.84	0.41	0.50
JMMD-HSIC	0.61	0.28	0.27	0.36	0.45	0.44	0.47	0.35	0.84	0.40	0.45
CoDATS	0.66	0.44	0.38	0.38	0.74	0.57	0.56	0.33	0.83	0.32	0.52
RAINCOAT	0.67	0.29	0.26	0.32	0.74	0.53	0.58	0.40	0.87	0.42	0.50
AdvSKM	0.61	0.35	0.32	0.38	0.46	0.42	0.40	0.34	0.86	0.35	0.45
CALDA	0.64	0.31	0.37	0.40	0.63	0.77	0.50	0.39	0.85	0.42	0.53
CLUDA	0.68	0.36	0.28	0.44	0.70	0.58	0.49	0.37	0.86	0.47	0.52
CoTMix	0.64	0.27	0.21	0.26	0.59	0.29	0.44	0.39	0.89	0.44	0.44
ConFGD	0.58	0.32	0.31	0.39	0.41	0.33	0.32	0.35	0.82	0.42	0.43
ACON	0.65	0.51	0.39	0.44	0.74	0.79	0.85	0.40	0.87	0.48	0.60
GLA-DA	0.49	0.23	0.21	0.23	0.69	0.42	0.43	0.37	0.88	0.46	0.44
EDEN	0.77	0.49	0.52	**0.54**	0.78	0.59	0.60	**0.47**	0.91	**0.57**	0.62
META	**0.79**	**0.54**	**0.53**	**0.54**	**0.80**	**0.84**	**0.86**	0.41	**0.97**	0.46	**0.67**

Table 18: Macro-F1 Score of each method on WISDM across 10 domain transfer scenarios.

Method	$2 \to 32$	$4 \to 15$	$7 \to 30$	$12 \to 7$	$12 \to 19$	$18 \to 20$	$20 \to 30$	$21 \to 31$	$25 \to 29$	$26 \to 2$	Avg
Source	0.64	0.51	0.38	0.48	0.47	0.54	0.41	0.17	0.30	0.54	0.44
CDAN	0.76	0.33	0.73	0.61	0.36	0.57	0.67	0.28	0.17	0.63	0.51
HoMM	0.55	0.38	0.58	0.45	0.36	0.60	0.55	0.29	0.28	0.53	0.46
MMDA	0.57	0.39	0.72	0.44	0.56	0.52	0.56	0.28	0.29	0.64	0.50
DSAN	0.54	0.44	0.62	0.53	0.51	0.77	0.59	0.33	0.26	0.74	0.53
JMMD-HSIC	0.70	0.30	0.74	0.62	0.36	0.62	0.63	0.32	0.23	0.64	0.52
CoDATS	0.63	0.37	0.83	0.39	0.45	0.76	0.61	0.27	0.27	0.70	0.53
RAINCOAT	0.77	0.70	0.72	0.69	0.62	0.85	0.74	0.61	0.52	0.82	0.70
AdvSKM	0.58	0.50	0.59	0.52	0.52	0.63	0.50	0.29	0.27	0.48	0.49
CALDA	0.72	0.49	0.80	0.65	0.51	0.78	0.63	0.30	0.13	0.69	0.57
CLUDA	0.60	0.49	0.73	0.73	0.39	0.81	0.63	0.30	0.27	0.56	0.55
CoTMix	0.69	0.74	0.80	0.67	0.52	0.41	0.58	0.24	0.20	0.71	0.56
ConFGD	0.73	0.76	0.86	0.71	0.69	0.83	0.46	0.33	0.29	0.72	0.64
ACON	0.91	0.62	0.95	0.58	0.52	0.87	0.76	**0.71**	**0.62**	0.82	0.74
GLA-DA	0.42	0.53	0.53	0.44	0.50	0.52	0.52	0.26	0.24	0.60	0.46
EDEN	0.69	**0.97**	0.99	0.83	0.70	0.82	0.78	0.44	0.21	**1.00**	0.74
META	**0.92**	0.93	**1.00**	**0.95**	**0.80**	**0.90**	**0.93**	0.48	0.33	**1.00**	**0.82**

Table 19: Macro-F1 Score of each method on HAR across 10 domain transfer scenarios.

Method	$2 \to 11$	$6 \to 23$	$7 \to 13$	$9 \to 18$	$12 \to 16$	$13 \to 19$	$18 \to 21$	$20 \to 6$	$23 \to 13$	$24 \to 12$	Avg
Source	0.78	0.72	0.83	0.73	0.68	0.81	0.98	0.81	0.71	0.89	0.79
CDAN	**1.00**	0.96	0.90	0.80	0.84	**1.00**	**1.00**	**1.00**	0.93	0.99	0.94
HoMM	**1.00**	0.87	0.90	0.75	0.77	0.98	0.99	0.87	0.85	0.97	0.90
MMDA	**1.00**	0.96	0.87	0.61	0.66	**1.00**	**1.00**	**1.00**	0.75	0.98	0.88
DSAN	0.90	0.68	0.73	0.60	0.46	0.98	**1.00**	0.86	0.65	0.97	0.78
JMMD-HSIC	0.99	0.87	0.91	0.76	0.67	0.98	**1.00**	0.90	0.80	0.98	0.89
CoDATS	**1.00**	0.96	0.92	0.79	0.67	**1.00**	**1.00**	0.99	0.86	**1.00**	0.92
RAINCOAT	**1.00**	0.96	0.91	0.81	0.81	**1.00**	**1.00**	0.99	0.91	**1.00**	0.94
AdvSKM	0.99	0.85	0.91	0.71	0.65	0.95	**1.00**	0.83	0.82	**1.00**	0.87
CALDA	0.97	0.91	0.93	0.73	0.81	**1.00**	**1.00**	0.99	0.92	0.98	0.92
CLUDA	**1.00**	0.96	0.93	0.81	0.69	**1.00**	**1.00**	**1.00**	0.89	**1.00**	0.93
CoTMix	0.64	0.52	0.89	0.64	0.84	0.80	0.98	0.92	0.86	0.89	0.80
ConFGD	**1.00**	0.91	0.95	**0.96**	0.89	**1.00**	**1.00**	**1.00**	0.90	**1.00**	0.96
ACON	**1.00**	0.97	0.97	0.90	0.83	**1.00**	**1.00**	**1.00**	0.91	**1.00**	0.96
GLA-DA	0.98	0.87	0.91	0.75	0.69	0.98	**1.00**	0.90	0.79	0.98	0.89
EDEN	**1.00**	0.98	0.99	0.94	0.87	**1.00**	**1.00**	**1.00**	0.99	**1.00**	0.98
META	**1.00**	**1.00**	**1.00**	**0.96**	**0.93**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**0.99**

Table 20: Macro-F1 Score of each method on Sleep-EDF across 10 domain transfer scenarios.

Method	$9 \to 11$	$2 \to 5$	$12 \to 5$	$7 \to 18$	$16 \to 1$	$9 \to 14$	$4 \to 12$	$10 \to 7$	$6 \to 3$	$8 \to 10$	Avg
Source	0.39	0.49	0.20	0.57	0.52	0.33	0.34	0.31	0.45	0.21	0.38
CDAN	0.52	0.63	0.66	0.66	0.64	0.69	0.54	0.75	0.76	0.55	0.64
HoMM	0.55	0.61	**0.70**	0.62	0.63	**0.73**	0.51	0.75	0.76	0.51	0.64
MMDA	0.39	0.63	0.68	0.69	0.63	0.70	0.55	0.75	**0.77**	0.51	0.63
DSAN	0.43	0.50	0.69	0.63	0.58	0.68	0.56	0.56	0.61	0.25	0.55
JMMD-HSIC	0.53	0.61	0.69	0.72	0.62	0.71	0.53	0.72	**0.77**	0.46	0.63
CoDATS	0.39	0.54	0.64	0.72	0.63	0.59	0.62	0.52	0.51	0.33	0.55
RAINCOAT	0.53	0.65	0.69	0.71	0.64	0.70	0.62	0.67	0.67	0.38	0.63
AdvSKM	0.41	0.61	0.68	0.56	0.59	0.66	0.58	0.74	0.74	0.42	0.60
CALDA	0.58	0.55	0.67	0.69	0.63	0.70	0.59	0.66	**0.77**	0.39	0.62
CLUDA	0.50	0.65	0.68	0.65	0.63	0.68	0.55	0.64	0.76	0.46	0.62
CoTMix	0.48	0.60	0.53	0.73	0.64	0.68	0.50	0.52	0.71	0.47	0.59
ConFGD	0.27	0.31	0.25	0.28	0.19	0.48	0.26	0.30	0.53	0.20	0.31
ACON	0.34	0.54	0.59	0.65	0.62	0.55	0.37	0.47	0.56	0.43	0.51
GLA-DA	0.52	0.57	0.64	0.59	0.61	0.69	0.51	0.69	0.76	0.52	0.61
EDEN	0.54	0.49	0.58	0.65	0.63	0.64	0.44	0.44	0.60	0.23	0.52
META	**0.65**	**0.73**	**0.70**	**0.80**	**0.77**	**0.73**	**0.70**	**0.78**	0.74	**0.60**	**0.72**

Table 21: Macro-F1 Score of each method on CAP across 10 domain transfer scenarios.

Method	$0\to1$	$0\to3$	$0\to4$	$1\to0$	$1\to4$	$2\to3$	$3\to0$	$3\to1$	$4\to1$	$4\to3$	Avg
Source	0.37	0.71	0.63	0.45	0.48	0.42	0.63	0.43	0.37	0.67	0.52
CDAN	0.50	0.63	0.49	0.56	0.55	0.45	0.65	0.55	0.50	0.67	0.55
HoMM	0.56	0.72	0.67	0.55	0.59	0.52	0.67	0.44	0.39	0.68	0.58
MMDA	0.36	**0.74**	0.65	0.52	0.59	0.53	0.64	0.27	0.43	0.65	0.54
DSAN	0.57	0.60	0.59	0.56	0.58	0.37	0.60	0.47	0.41	0.68	0.54
JMMD-HSIC	0.55	0.73	0.63	0.48	0.52	0.50	0.64	0.47	0.41	0.68	0.56
CoDATS	0.58	0.72	0.66	0.58	0.62	0.43	**0.69**	0.56	0.51	0.70	0.61
RAINCOAT	0.50	0.64	0.54	0.48	0.54	0.34	0.54	0.58	0.59	0.51	0.53
AdvSKM	0.53	0.72	0.59	0.48	0.47	0.50	0.66	0.44	0.43	0.68	0.55
CALDA	0.55	0.61	0.60	0.53	0.58	0.37	0.67	0.55	0.54	0.67	0.57
CLUDA	0.51	0.57	0.55	0.56	0.60	0.37	0.62	0.42	0.40	0.55	0.51
CoTMix	0.43	0.66	0.62	0.34	0.27	0.27	0.56	0.43	0.39	0.62	0.46
ConFGD	0.40	0.72	0.59	0.43	0.49	**0.55**	0.63	0.33	0.36	0.70	0.52
ACON	0.63	0.67	0.62	0.60	**0.65**	0.52	0.67	0.70	0.61	0.69	0.64
GLA-DA	0.53	0.73	0.67	0.45	0.48	0.49	0.64	0.49	0.41	0.69	0.56
EDEN	0.32	0.60	0.55	0.47	0.49	0.37	0.60	0.37	0.40	0.52	0.47
META	**0.67**	**0.74**	**0.70**	**0.61**	0.62	0.54	0.68	**0.72**	**0.67**	**0.72**	**0.67**

Table 22: Macro-F1 Score of each method on EMG across 10 domain transfer scenarios.

Method	$0\to1$	$0\to2$	$0\to3$	$1\to2$	$1\to3$	$2\to0$	$2\to1$	$2\to3$	$3\to1$	$3\to2$	Avg
Source	0.85	0.70	0.72	0.68	0.72	0.64	0.80	0.71	0.81	0.71	0.73
CDAN	0.85	0.74	0.77	0.78	0.79	0.70	0.82	0.80	0.81	0.73	0.78
HoMM	0.82	0.73	0.80	0.78	0.79	0.71	0.83	0.76	0.80	0.75	0.78
MMDA	0.83	0.73	0.79	0.77	0.79	0.74	0.85	0.76	0.81	0.75	0.78
DSAN	0.86	0.75	0.79	0.76	0.77	0.71	0.83	0.80	0.82	0.74	0.78
JMMD-HSIC	0.82	0.72	0.76	0.77	0.78	0.72	0.81	0.74	0.81	0.74	0.77
CoDATS	0.86	0.74	0.76	0.78	0.81	0.66	0.84	0.76	0.79	0.76	0.78
RAINCOAT	0.84	0.74	0.76	0.78	0.80	0.68	0.84	0.77	0.82	0.76	0.78
AdvSKM	0.79	0.71	0.74	0.78	0.81	0.61	0.75	0.69	0.79	0.70	0.74
CALDA	0.69	0.69	0.62	0.67	0.63	0.55	0.62	0.63	0.73	0.64	0.65
CLUDA	0.84	0.75	0.73	0.78	0.79	0.71	0.83	0.77	0.81	0.76	0.78
CoTMix	0.79	0.74	0.63	0.76	0.79	0.70	0.79	0.73	0.76	0.75	0.74
ConFGD	0.84	0.73	0.74	0.76	0.81	0.73	0.82	0.77	0.80	0.75	0.77
ACON	**0.92**	**0.81**	0.81	**0.82**	0.85	0.72	0.90	**0.82**	0.85	**0.81**	0.83
GLA-DA	0.81	0.72	0.73	0.77	0.78	0.71	0.80	0.72	0.81	0.73	0.76
EDEN	0.88	0.78	0.79	**0.82**	0.84	0.80	**0.91**	0.80	0.84	0.75	0.82
META	0.91	**0.81**	**0.84**	**0.82**	**0.86**	**0.81**	**0.91**	0.80	**0.87**	0.80	**0.84**

Table 23: Macro-F1 Score of each method on PCL across 6 domain transfer scenarios.

Method	$0 \to 1$	$0 \to 2$	$1 \to 0$	$1 \to 2$	$2 \to 0$	$2 \to 1$	Avg
Source	0.59	0.57	0.58	0.60	0.56	0.67	0.59
CDAN	0.67	0.59	0.63	0.61	0.54	0.57	0.60
HoMM	0.69	0.57	0.61	0.64	0.59	0.68	0.63
MMDA	0.35	0.36	0.57	0.35	0.36	0.47	0.41
DSAN	0.68	0.57	0.63	0.61	0.57	0.66	0.62
JMMD-HSIC	0.68	**0.61**	0.62	0.62	0.57	0.69	0.63
CoDATS	0.68	0.60	0.62	0.58	0.59	**0.70**	0.63
RAINCOAT	0.57	0.56	0.62	0.62	0.59	0.69	0.61
AdvSKM	0.69	0.60	0.63	0.63	**0.60**	0.69	**0.64**
CALDA	0.56	0.53	0.63	0.54	0.39	0.41	0.51
CLUDA	**0.70**	0.34	0.62	0.64	0.52	0.69	0.58
CoTMix	0.50	0.34	**0.65**	0.40	0.33	0.33	0.43
ConFGD	0.68	0.59	0.62	0.62	0.57	0.69	0.63
ACON	0.68	0.59	0.61	0.62	**0.60**	0.67	0.63
GLA-DA	0.67	0.59	0.61	0.61	0.57	0.69	0.62
EDEN	0.48	0.51	0.60	0.61	0.55	0.34	0.52
META	0.69	0.58	0.64	**0.66**	0.59	0.68	**0.64**