RETHINKING SPATIAL ALIGNMENT IN UDA FOR WEARABLE MULTIVARIATE TIME SERIES: A DIAGNOSTIC ANALYSIS AND A SIMPLE REMEDY

Anonymous authors
Paper under double-blind review

ABSTRACT

Wearable sensing models often fail when deployed on new users, devices, or sensor placements, because the meaning of each sensor channel is inconsistent across these settings. Unsupervised domain adaptation (UDA) for multivariate time series (MTS) aims to close this gap, yet existing methods mainly address temporal distribution shift or align domains using spatial structure learned within each domain separately. Through a diagnostic analysis, we identify a distinct failure mode in **wearable** MTS, which we term *sensor-wise domain shift*. Current alignment strategies do not model this cross-domain channel mismatch explicitly and therefore struggle to correct it. We propose **IDSA** (***I**nter-**D**omain **S**ensor **A**lignment*), a backbone-agnostic plug-in module that aligns sensors explicitly across domains. IDSA combines (i) *inter-domain sensor transport*, which learns a source–target sensor correspondence matrix from domain-specific sensor embeddings and maps target channels into the source sensor space, and (ii) *channel decorrelation*, a regularizer that suppresses redundant or noisy channel couplings and prevents representation collapse. The resulting spatial transportation loss can be interpreted as a sensor-level discrete optimal-transport objective. IDSA improves average activity-classification performance over recent baselines on HAR and sEMG benchmarks across most evaluated scenarios.

1 INTRODUCTION

Wearable sensing underpins applications such as health monitoring (Inoue et al., 2019), rehabilitation (Schrader et al., 2020), and human-computer interaction (Qi et al., 2019), through tasks including human activity recognition (Zhong et al., 2022) and gesture classification (Tchantchane et al., 2023). A body-worn device records several signals at once, such as the axes of accelerometers and gyroscopes or the outputs of electrodes placed over different muscles. Each recording can therefore be treated as a multivariate time series (MTS), with every signal forming one channel. The usable information lies both in how each channel evolves over time and in how the channels relate to one another, and both change across users, devices, and re-wearings. Unsupervised domain adaptation (UDA), which uses labeled source data together with unlabeled target data, is the standard tool for mitigating such shifts (Huang et al., 2022; Ozyurt et al., 2023).

Most MTS-UDA methods share an assumption that is rarely stated: the channel index is a persistent physical identifier. In many multivariate time-series settings, such as plant instrumentation or grid metering, this holds: channel i is the same transducer at the same location in every domain. Wearable data violate it silently: re-mounting a device, or handing it to another person, changes the mapping between physical signal sources and recorded channels. When a circular surface electromyography (sEMG) armband is reattached rotated by $30°$, the same muscle is read by a different electrode; when an inertial unit is placed in a different orientation, its axes mix. Channel i in the target domain then no longer measures what channel i measured in the source domain, and the cross-channel dependency structure changes with it. We call this *sensor-wise domain shift*: a shift in the semantics of the channels themselves, as opposed to a shift in temporal dynamics or in the activities performed.

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.

Existing MTS-UDA methods do not model this shift. Most address distribution shift along the temporal dimension (He et al., 2023; Liu et al., 2024); those that exploit spatial structure capture it *within* each domain, through structure-aware encoders such as GNNs (Lai et al., 2021) or, as in SEA (Wang et al., 2023) and SASA (Cai et al., 2021), by aligning sensor graphs or sparse associative structures learned separately in each domain. Both families rely on positional correspondence that channel i plays the same role across domains, which sensor shift breaks and within-domain alignment cannot restore. Wearable literature counters sensor displacement, orientation change (Kunze & Lukowicz, 2008; Banos et al., 2014), and electrode shift (Young et al., 2011) using orientation-invariant preprocessing (Yurtman & Barshan, 2017) or cross-session re-normalization (Du et al., 2017). However, these build pre-deployment invariance while assuming pre-aligned channels; to our knowledge, none estimate the actual source–target channel correspondence encountered in the target domain.

These observations motivate us to rethink spatial alignment in UDA for wearable MTS. Our contributions are:
(1) Diagnostic analysis: Under a latent channel-mixing model, we show how sensor-wise domain shift induces mismatched feature and structural representations that may not be resolved by structure-aware feature alignment or structure-driven alignment. We observe real-world sensor-wise variability consistent with this analysis, demonstrate the failure mode under progressively stronger controlled sensor-wise shifts, and show that an explicit source–target channel map can mitigate it.
(2) Method: We propose **IDSA (Inter-Domain Sensor Alignment)**, a backbone-agnostic *plug-in* that learns a doubly stochastic target-to-source transport matrix, i.e., a nonnegative re-mixing of channels covering permutations and soft permutations such as electrode-ring rotations, and re-expresses target channels in the source channel space before any downstream UDA objective. The transport admits a sensor-level optimal-transport interpretation; a *channel decorrelation* loss keeps transported channels distinguishable and prevents representation collapse.
(3) Evaluation: On four wearable datasets (five benchmark settings) spanning cross-subject HAR and cross-repetition sEMG gesture recognition, IDSA improves the average performance of all four UDA methods it is plugged into, is competitive or superior in most scenarios, and is substantially more robust than strong baselines under controlled channel permutation. We also benchmark against source-supervised foundation models on HAR to assess the practical relevance of sensor-wise adaptation alongside pretrained alternatives.

2 Related Work

Wearable sensor data for activity (Roggen et al., 2010) and gesture recognition (Shen et al., 2019) vary across subjects and wearing conditions (Wang et al., 2023), and unsupervised domain adaptation (UDA) addresses this by aligning source and target features adversarially (Ganin et al., 2016; Long et al., 2018) or statistically (Sun & Saenko, 2016; Redko et al., 2019). Applied to multivariate time series (MTS), such methods treat channels as fixed input dimensions and thus assume positional correspondence, i.e., that channel i measures the same quantity in both domains.

MTS-specific UDA additionally exploits spatial structure (Wang et al., 2023; Lu & Sun, 2024; Guo et al., 2025): structure-aware approaches encode within-domain sensor dependencies, e.g., with GNNs (Lai et al., 2021), before aligning features, while structure-driven approaches align the structures themselves, e.g., sensor graphs in SEA (Wang et al., 2023) and sparse associative or causal structures in SASA (Cai et al., 2021) and its causal extension (Li et al., 2024). In all cases the structure is learned separately in each domain and matched by channel index, presuming a known and stable source–target sensor correspondence; Section 4 shows that sensor-wise domain shift breaks this presumption and that within-domain structure alignment cannot repair it.

Optimal transport (OT) for domain adaptation couples source and target samples in a shared feature space (Courty et al., 2017; Kerdoncuff et al., 2020) or a latent time–frequency space (He et al., 2023). IDSA instead transports sensor channels, with a coupling learned end-to-end under a downstream UDA objective, to re-express target channels in the source channel space before adaptation (Appendix A.4.2).

Time-series foundation models (Zhang et al., 2024; Feofanov et al., 2026; Luo et al., 2026) pursue robustness through large-scale pretraining rather than per-domain adaptation, so their robustness depends on whether the pretraining data and augmentations cover the shift met at test time. IDSA instead estimates the source–target sensor correspondence per domain pair from unlabeled target data, which matters when the deployed sensor layout differs from the pretraining configuration, as with custom electrode arrays such as SD-gesture (Lee et al., 2023).

3 PROBLEM SETUP AND FORMULATION

We consider a labeled source dataset $\mathcal{D}_s = \{(\mathbf{X}_s^i, y_s^i)\}_{i=1}^{n_s}$ and an unlabeled target dataset $\mathcal{D}_t = \{\mathbf{X}_t^i\}_{i=1}^{n_t}$, where n_s and n_t denote the numbers of source and target samples, respectively. Each instance $\mathbf{X}^i \in \mathbb{R}^{N\times T}$ is a multivariate sensor sequence, where N denotes the number of sensor-channels and T represents the number of time steps.

Throughout the paper, we use the term *sensor-channel* to denote each input variable of the model. A physical sensor device may contribute multiple sensor-channels; for example, an IMU provides six channels, three from the accelerometer and three from the gyroscope. For clarity, we omit the sample superscript i in the following discussion. Thus, the source and target data instances will be denoted simply as $\mathbf{X}_s$ and $\mathbf{X}_t$, respectively.

4 DIAGNOSTIC ANALYSIS

4.1 THEORETICAL ANALYSIS: ASSUMPTION GAP AND SENSOR-WISE DOMAIN SHIFT

Following the standard domain adaptation theory of Ben-David et al. (2010), the target domain error, $\epsilon_T(h)$, is bounded by:

$$\epsilon_T(h) \le \underbrace{\epsilon_S(h)}_{(1)} + \underbrace{\frac{1}{2}\text{dist}(S,T)}_{(2)} + \underbrace{\lambda}_{(3)}, \tag{1}$$

where $\epsilon_S(h)$ is the source risk, $\text{dist}(S,T)$ is the domain discrepancy, and $\lambda = \min_{h^*}(\epsilon_S(h^*) + \epsilon_T(h^*))$ is the ideal joint error of an optimal classifier h^*. Many UDA objectives focus on minimizing the first two terms, implicitly assuming that λ is small, i.e., that a *single* classifier can perform well on both domains.

4.1.1 ROLE OF λ IN WEARABLE MTS CLASSIFICATION

We assume sensor observations across both domains derive from a shared latent canonical signal $\mathbf{U} \in \mathbb{R}^{N\times T}$. This formalizes the core UDA invariance assumption: without a shared latent representation, the ideal joint error λ in Equation 1 can remain large despite marginal alignment. Under this assumption, wearable sensor signals $\mathbf{X}_d$ are modeled as:

$$\mathbf{X}_d = \mathbf{M}_d\mathbf{U} + \xi_d, \qquad d \in \{s,t\}, \tag{2}$$

where $\mathbf{M}_d \in \mathbb{R}^{N\times N}$ encodes sensor configuration differences (e.g., rotation, scaling, permutation), ξ_d is noise, and s,t denote source and target domains. For wearable sensors, this is physically intuitive: reorienting a device rotates tri-axial readings of the same motion.

Below, we illustrate how sensor-wise domain shift can challenge the small-joint-error assumption for two representative UDA families that exploit structural information.

Structure-aware classic UDA. These methods use an encoder ϕ capturing within-domain cross-variable dependency Σ_d to align extracted *features* ($\phi(\mathbf{X}_s)$ and $\phi(\mathbf{X}_t)$). For example, a GCN (Kipf & Welling, 2017) with a correlation-based adjacency matrix yields a source embedding $\mathbf{H}_s := \phi(\mathbf{X}_s) \approx (\mathbf{M}_s\mathbf{U}\mathbf{U}^\top\mathbf{M}_s^\top)\mathbf{M}_s\mathbf{U}$ governed by $\mathbf{M}_s$, where parentheses denote learned graph topology. Because target embedding $\mathbf{H}_t$ depends similarly on $\mathbf{M}_t$, $\mathbf{M}_s \neq \mathbf{M}_t$ can lead to embeddings in mismatched sensor-coordinate systems, i.e., $\mathbf{H}_s \neq \mathbf{H}_t$. Thus, a classifier optimized on $\mathbf{H}_s$ may fail to generalize to $\mathbf{H}_t$. Consequently, feature alignment alone does not guarantee the existence of a shared low-error predictor required for a small λ.

Structure-driven UDA. These methods assume domain-invariant variable dependencies: a shared spatial extractor f defines a structure $\mathbf{S}_d := f(\Sigma_d)$ (associative, causal, or graphical) for each domain $d \in \{s,t\}$, and the domains are aligned by enforcing $\mathbf{S}_s \approx \mathbf{S}_t$.

Under the signal model in Equation 2, cross-channel second-order structure transforms as

$$\Sigma_d = \text{Cov}(\mathbf{X}_d) \approx \mathbf{M}_d\,\text{Cov}(\mathbf{U})\,\mathbf{M}_d^\top. \tag{3}$$

This shows that when $\mathbf{M}_s \neq \mathbf{M}_t$, Σ_s and Σ_t can differ substantially despite invariant $\text{Cov}(\mathbf{U})$. Consequently, forcing f to satisfy $\mathbf{S}_s \approx \mathbf{S}_t$ may not be sufficient for reducing λ in Equation 1.

Although our derivation uses a linear mixing model $\mathbf{X}_d = \mathbf{M}_d\mathbf{U} + \xi_d$ for tractability, replacing $\mathbf{M}_d$ with a smooth nonlinear map g_d can exhibit the same intuition: domain-specific channel semantics induce a mismatch in Σ_d that may not be fully resolved by within-domain encoding alone. In this case, the ideal inter-domain channel map can be viewed as a nonlinear transformation $g_s \circ g_t^{-1}$.

4.1.2 SENSOR-WISE DOMAIN SHIFT AND SIMPLE REMEDY

We formalize *sensor-wise domain shift* in wearable MTS, a practically relevant form of domain shift in which the semantics of observed variables (sensor-channels) can change across domains, thereby potentially violating the assumption that the ideal joint error λ in Equation 1 remains small.

Definition 1 (Sensor-wise domain shift)**.** ***Sensor-wise domain shift*** *in wearable multivariate time series refers to the distributional discrepancy induced by domain-dependent sensor semantics ($\mathbf{M}_d$), such as differences in sensor placement or orientation across domains.*

Under Equation 2, a linear classifier trained on $\mathbf{X}_s$ learns decision directions in the source sensor-coordinate system; applied to $\mathbf{X}_t$, it interacts with the target mixing matrix, leading to a mismatch that can be expressed through $\mathbf{M}_s^{-1}\mathbf{M}_t$ when source-domain coordinates are used. We define a cross-domain channel mapping $\mathbf{A}_{s\leftarrow t}$ such that the transformed target channels become comparable to the source:

$$\mathbf{X}_t^* = \mathbf{A}_{s\leftarrow t}\mathbf{X}_t, \qquad \Sigma_s \approx \mathbf{A}_{s\leftarrow t}\,\Sigma_t\,\mathbf{A}_{s\leftarrow t}^\top. \tag{4}$$

Under the mixing model in Equation 2 (and ignoring noise), one valid choice is $\mathbf{A}_{s\leftarrow t} = \mathbf{M}_s\mathbf{M}_t^{-1}$, when $\mathbf{M}_t$ is invertible, which "undoes" the domain-specific channel mixing. This motivates explicitly modeling source–target variable correspondence to address sensor-wise domain shift.

4.2 EMPIRICAL ANALYSIS

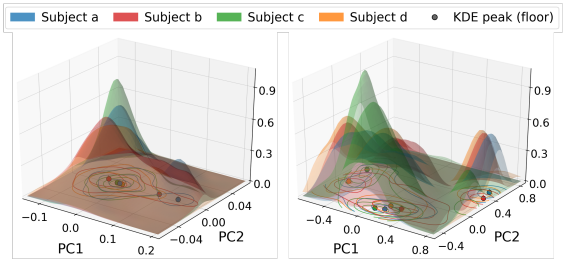
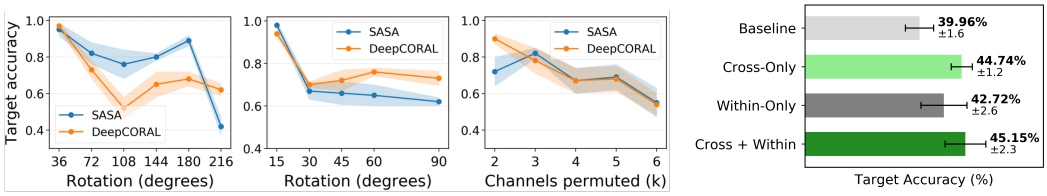


Figure 1: (Left) Temporal profiles and (Right) sensor-wise profiles. KDE surfaces and floor contours show domain-wise distributions in the respective PCA spaces. Markers indicate prominent local density peaks.

Real-world Observation of Domain Variability. To examine whether the sensor-wise domain shift occurs in real data, we analyze temporal and sensor-wise profiles from the Opportunity HAR dataset (Roggen et al., 2010). Per window, temporal profiles average absolute signal magnitude across channels over time, while sensor-space profiles compute channel-wise RMS activation. Applying PCA separately to both, we estimate 2D domain-wise densities via KDE, visualized as 3D surfaces with floor contours. Fig. 1 (Left) shows temporal profiles vary along PC1 while high-density regions remain concentrated along PC2. Conversely, Fig. 1 (Right) reveals distinct, dynamic sensor-wise patterns, with high-density peaks shifted across both PC1 and PC2. This observation suggests that sensor-wise profiles capture cross-domain variability less apparent in temporal views, motivating controlled evaluations of UDA robustness to sensor-wise shifts.

Controlled Analysis of Sensor-wise Domain Shift. We stress-test representative *structure-aware* UDA (feature alignment via spatial encoder; DeepCORAL (Sun & Saenko, 2016)) and *structure-driven* UDA (explicit structure alignment; SASA (Cai et al., 2021)) baselines under increasing shift severity across three stylized synthetic settings: **Cross-subject sensor orientation shift**, **Cross-repetition sensor placement shift**, and **Variable permutation**. Figure 2(a) shows that while both baselines perform well under mild shifts, their accuracy degrades as shift severity increases (left to right). This trend is consistent with our analysis: when source–target sensor correspondence becomes increasingly unstable, methods relying primarily on feature alignment or invariant within-domain structure may fail to reduce $\epsilon_T(h)$.

Effect of Cross-domain Sensor Correspondence. We further validate our findings on real-world data using 10 randomly selected sensor-channels from two subjects in the Opportunity HAR dataset (Roggen et al., 2010). Figure 2(b) shows that applying correspondence-based target transformation improves target accuracy in this diagnostic setup (CROSS-ONLY > BASELINE;



(a) (b)

Figure 2: (a) Three plots showing target performance under controlled sensor-wise domain shift scenarios: orientation shift, placement shift, and variable permutation from left to right. (b) Empirical diagnosis validating the effectiveness of explicitly modeling inter-domain sensor relations.

CROSS+WITHIN > WITHIN-ONLY). This supports the need to estimate source–target variable correspondence in a data-dependent manner for each domain pair, rather than relying solely on within-domain sensor relations. Additional details are in the Appendix.

5 METHODOLOGY

This section introduces ***I**nter-**D**omain **S**ensor **A**lignment* (**IDSA**), a plug-in module designed to mitigate *sensor-wise domain shift* in existing MTS-UDA methods. Figure 3 illustrates the overall framework.

5.1 INTER-DOMAIN SENSOR TRANSPORTATION

Motivated by the simple remedy introduced in Section 4, namely the cross-domain mapping $\mathbf{A}_{s\leftarrow t}$ that explicitly models source–target sensor correspondence, we operationalize this idea through a learnable inter-domain sensor relation. Specifically, we introduce domain-specific source and target sensor embeddings, $\mathbf{P}_s, \mathbf{P}_t \in \mathbb{R}^{N\times T}$, and define the approximated inter-domain sensor relation $\mathbf{A}_{st} \in \mathbb{R}^{N\times N}$ as:

$$\mathbf{A}_{st} = \texttt{Norm}(\sigma(\mathbf{P}_s\mathbf{P}_t^\top)), \tag{5}$$

where σ is an activation function, and $\texttt{Norm}$ is a normalization function that makes the matrix approximately doubly stochastic. This is implemented via Sinkhorn–Knopp normalization (Knight, 2008).

Spatial transportation loss The decomposition in Equation 2 is inherently non-unique. To select a decomposition well-suited for cross-domain alignment, we constrain $\mathbf{A}_{st}$ using a pairwise source–target sensor-channel cost, motivated by the assumption that channels with similar observed signals are more likely to share related latent dynamics:

$$\mathcal{L}_{st} = \langle \mathbf{A}_{st}, \mathbf{C}\rangle = \sum_{i=1}^{N}\sum_{j=1}^{N}(\mathbf{A}_{st})_{ij}\mathbf{C}_{ij}, \qquad \mathbf{C}_{ij} = \|\mathbf{X}_s(i,\cdot) - \mathbf{X}_t(j,\cdot)\|_2^2, \tag{6}$$

where $\langle\cdot,\cdot\rangle$ denotes the Frobenius inner product, and $\mathbf{C} \in \mathbb{R}^{N\times N}$ is the ground-cost matrix of squared Euclidean distances between source and target sensor-channel profiles. For mini-batch training, these profiles are averaged separately over the batch dimension within each domain, requiring no source–target window pairing. The sensor-wise distance can also be extended to other ground costs, such as cosine or correlation-based distances.

5.2 INTEGRATION OF INTER- AND WITHIN-DOMAIN SENSOR RELATION

We enrich input data with domain-specific sensor-channel information using learnable sensor embeddings $\mathbf{P}_s$ and $\mathbf{P}_t$:

$$\tilde{\mathbf{X}}_s = \mathbf{X}_s + \mathbf{P}_s, \qquad \tilde{\mathbf{X}}_t = \mathbf{X}_t + \mathbf{P}_t. \tag{7}$$

We then apply the inter-domain relation matrix $\mathbf{A}_{st}$ as a transport map to align target sensor arrangement with the source:

$$\hat{\mathbf{X}}_t = \mathbf{A}_{st}\,\tilde{\mathbf{X}}_t. \tag{8}$$

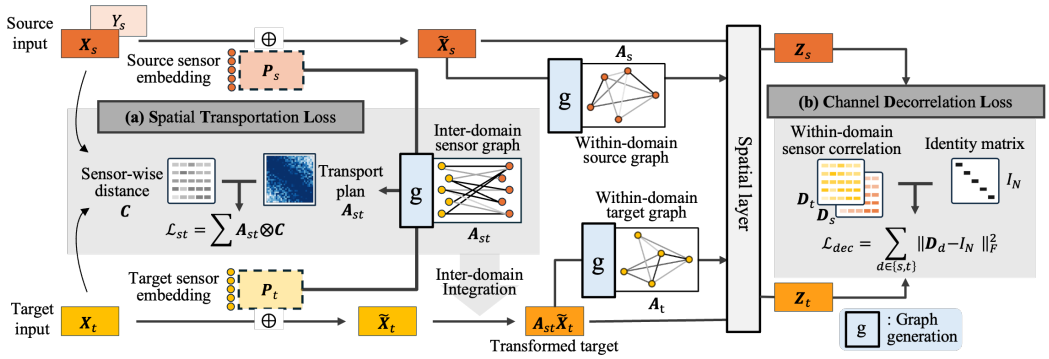


Figure 3: Illustration of IDSA. IDSA learns domain-specific sensor embeddings ($\mathbf{P}_s, \mathbf{P}_t$) and is trained with a spatial transportation loss ($\mathcal{L}_{\text{st}}$) and a channel decorrelation loss ($\mathcal{L}_{\text{dec}}$) to align target sensor-channels with the source while preserving within-domain channel distinctiveness.

This transformation maps target sensor-channels into the source sensor-coordinate space and is intended to reduce sensor-wise domain shift as well as the source–target distribution discrepancy.

To capture within-domain spatial relationships, we model the N sensor channels as graph nodes using a GNN (Kipf & Welling, 2017), with domain adjacency matrices $\mathbf{A}_s = \sigma(\tilde{\mathbf{X}}_s \tilde{\mathbf{X}}_s^\top)$ and $\mathbf{A}_t = \sigma(\hat{\mathbf{X}}_t \hat{\mathbf{X}}_t^\top)$. Adopting the GNN, features $\mathbf{Z}_s$ and $\mathbf{Z}_t$ are computed as:

$$\mathbf{Z}_s = \mathbf{A}_s \tilde{\mathbf{X}}_s \mathbf{W}_l, \quad \mathbf{Z}_t = \mathbf{A}_t \hat{\mathbf{X}}_t \mathbf{W}_l, \qquad \mathbf{W}_l \in \mathbb{R}^{T \times T},\ \mathbf{Z}_s, \mathbf{Z}_t \in \mathbb{R}^{N \times T}. \tag{9}$$

Channel decorrelation loss However, stacking message-passing layers increases node representation correlation. Recent work (Tai et al., 2026) shows that neighborhood aggregation amplifies representation interdependency, degrading generalization in graph domain adaptation. Because our sensor graph connects feature-similar sensors, this effect is particularly pronounced, motivating a decorrelation loss to penalize inter-sensor correlations:

$$\mathbf{D}_d = \frac{1}{T} \mathbf{Z}_d \mathbf{Z}_d^\top \in \mathbb{R}^{N \times N}, \qquad d \in \{s, t\}, \tag{10}$$

$$\mathcal{L}_{\text{dec}} = \sum_{d \in \{s,t\}} \left\| \mathbf{D}_d - \mathbf{I}_N \right\|_F^2, \tag{11}$$

where $\mathbf{I}_N$ is the $N \times N$ identity matrix and $\mathbf{D}_d$ denotes the channel-wise Gram matrix constructed from the spatial outputs for each domain d. The combined effect of $\mathcal{L}_{st}$ and $\mathcal{L}_{\text{dec}}$ is discussed from a regularization perspective in Section 5.4.

5.3 Training and Inference

Since IDSA is designed as a plug-in module that can be integrated with existing domain adaptation methods, a base domain adaptation loss $\mathcal{L}_{\text{da}}$ is also included. The final loss for IDSA is denoted as:

$$\mathcal{L} = \mathcal{L}_{\text{da}} + \lambda_1 \mathcal{L}_{\text{st}} + \lambda_2 \mathcal{L}_{\text{dec}}, \tag{12}$$

where λ_1 and λ_2 are hyperparameters that control the contributions of the proposed losses. At inference, we apply the learned target embedding $\mathbf{P}_t$ and transport $\mathbf{A}_{st}$ to held-out samples from the same unlabeled target domain used during adaptation; a new target domain requires re-estimating them.

5.4 Theoretical Analysis on IDSA

We formalize the spatial transportation procedure through an optimal transport (OT) formulation (Kolouri et al., 2017).

Proposition 1. *For a fixed ground-cost matrix* $\mathbf{C}$*, optimizing* $\mathcal{L}_{st}$ *over the full set of doubly stochastic matrices is equivalent, up to a factor of* N*, to a discrete 2-Wasserstein problem. If* $\mathbf{A}_{st}$ *achieves the global minimum of* $\mathcal{L}_{st}$*, it provides an optimal transport plan (up to constant scaling), so* $\min \mathcal{L}_{st} = N \cdot (W_2(\mathbf{X}_s, \mathbf{X}_t))^2$*, where* $W_2(\mathbf{X}_s, \mathbf{X}_t)$ *is the discrete 2-Wasserstein distance between the source and target sensor-channel sets.*

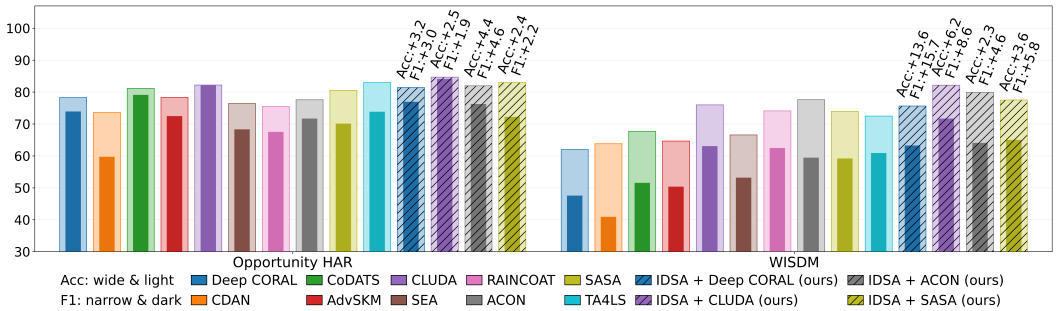


Figure 4: Cross-subject UDA performance on HAR datasets, averaged over all 12 scenarios for Opportunity HAR and the 10 selected scenarios for WISDM following Ozyurt et al. (2023). Wide/light bars denote accuracy, while narrow/dark bars denote F1-score. Hatched bars indicate IDSA-integrated variants. Numbers above the IDSA bars report percentage-point gains over the corresponding base models. Full per-scenario results are provided in Table 6 in the Appendix.

Proposition 1 thus characterizes $\mathcal{L}_{st}$ as a sensor-level discrete OT objective; its proof, an empirical check, and a comparison with instance-level OT baselines are given in Appendices A.2 and A.4.2.

Complementary roles of $\mathcal{L}_{\text{st}}$ and $\mathcal{L}_{\text{dec}}$. $\mathcal{L}_{st}$ directly biases $\mathbf{A}_{st}$ toward source–target channel pairs with low ground cost, while $\mathcal{L}_{dec}$ keeps transported channel representations *distinct*. $\mathcal{L}_{\text{st}}$ depends on the model only through $\mathbf{A}_{st}$, and its cost $\mathbf{C}$ comes from the raw signals. $\mathcal{L}_{\text{dec}}$ pushes $\mathbf{D}_d$ toward the identity, keeping channel embeddings nearly orthogonal after message passing. Because the spatial layer is permutation-equivariant, $\mathcal{L}_{\text{dec}}$ is invariant to reordering the transported channels and does not directly encode matching information, whereas $\mathcal{L}_{\text{st}}$ provides an explicit correspondence signal. Conversely, $\mathcal{L}_{\text{dec}}$ guards against a degenerate transport: a nearly uniform $\mathbf{A}_{st}$ replaces every target channel with the channel average, reducing channel-specific information. $\mathcal{L}_{\text{st}}$ penalizes this only when $\mathbf{C}$ clearly separates channel pairs, since a flat $\mathbf{C}$ gives all doubly stochastic matrices nearly equal cost. Thus, $\mathcal{L}_{\text{st}}$ and $\mathcal{L}_{\text{dec}}$ operate complementarily to ensure both accurate alignment and non-degenerate channel representations.

6 Experiment

6.1 Experimental Settings and Baselines

We conduct extensive evaluations to verify how IDSA performs compared to state-of-the-art methods on public wearable MTS datasets. The datasets are categorized into two groups: (1) human activity recognition (HAR) datasets for cross-subject adaptation, including the Opportunity HAR (Roggen et al., 2010) and the WISDM (Kwapisz et al., 2011), and (2) surface electromyography (sEMG) datasets for cross-repetition adaptation, including NinaPro DB5 (Pizzolato et al., 2017) and SD-gesture (Lee et al., 2023). For NinaPro DB5, we experiment on both the full 53-class setting, denoted as Nina5-53, and the 18-class subset, denoted as Nina5-18.

We compare IDSA against ten UDA baselines: Deep CORAL (Sun & Saenko, 2016), CDAN (Long et al., 2018), CoDATS (Wilson et al., 2020), AdvSKM (Liu & Xue, 2021), CLUDA (Ozyurt et al., 2023), SEA (Wang et al., 2023), RAINCOAT (He et al., 2023), ACON (Liu et al., 2024), SASA (Cai et al., 2021), and the recent plug-in method TA4LS (Na et al., 2025) (plugged into SASA for HAR datasets and Deep CORAL for sEMG). To evaluate IDSA's plug-in behavior across diverse adaptation objectives, we instantiate it with four representative baselines: *Deep CORAL* (statistical alignment), *CLUDA* (strong recent MTS-UDA), *ACON* (temporal-frequency adaptation), and *SASA* (structure-driven UDA). The remaining methods serve as non-plug-in comparison baselines. We additionally evaluate four pretrained time-series foundation-model configurations: MOMENT-small and MOMENT-large (Goswami et al., 2024), NormWear (Luo et al., 2026), and Mantis-8M (Feofanov et al., 2026). For a source-supervised comparison, each pretrained encoder is frozen, and a classifier is trained only on labeled source-domain data. Hyperparameter sensitivity, visualization, and additional ablation studies are in Appendix A.4.

6.2 UDA Performance Comparison

Cross-subject Figure 4 presents cross-subject performance (where source and target correspond to different subjects). IDSA improves average performance across the four plugged-in baselines

Table 1: Cross-repetition UDA performance on sEMG datasets. Results are reported as mean ± standard deviation over five shared random seeds. The **boldfaced** score denotes the best, and underlined represents the second-best result. Parentheses denote gains of IDSA over its base model.

Models	Nina5-18		Nina5-53		SD-gesture	
	Accuracy (%)	F1-score (%)	Accuracy (%)	F1-score (%)	Accuracy (%)	F1-score (%)
Deep CORAL	83.47 ±2.1	60.72 ±2.8	75.68 ±2.5	40.00 ±4.2	73.02 ±2.1	72.32 ±2.6
CDAN	83.48 ±1.9	61.11 ±2.2	75.82 ±3.1	39.75 ±5.7	72.84 ±2.5	71.87 ±2.1
CoDATS	81.74 ±1.6	57.43 ±3.0	67.58 ±2.0	15.74 ±3.3	75.48 ±2.9	74.90 ±2.3
AdvSKM	83.52 ±1.7	61.30 ±2.9	76.22 ±1.8	40.38 ±3.4	75.28 ±3.0	74.46 ±2.6
CLUDA	80.64 ±2.2	53.77 ±2.1	65.56 ±2.7	10.44 ±2.9	72.49 ±2.2	71.36 ±2.7
SEA	82.73 ±2.5	58.02 ±3.4	75.53 ±3.0	40.36 ±3.3	77.41 ±1.3	76.48 ±2.6
RAINCOAT	80.82 ±2.0	55.73 ±2.3	75.14 ±2.5	39.88 ±4.7	74.14 ±2.9	73.44 ±3.4
ACON	82.07 ±1.5	58.40 ±3.6	76.31 ±1.8	40.72 ±5.3	76.96 ±1.7	75.51 ±2.7
SASA	83.06 ±1.9	61.11 ±2.7	73.23 ±1.9	39.96 ±4.2	75.45 ±2.5	75.11 ±2.4
TA4LS	81.02 ±1.3	58.96 ±2.1	76.12 ±1.6	39.96 ±2.8	75.71 ±2.5	75.06 ±1.3
IDSA+Deep CORAL	**85.53** ±2.1 (+2.06)	**63.32** ±3.3 (+2.60)	**78.80** ±1.2 (+3.12)	**43.77** ±4.9 (+3.77)	80.15 ±2.1 (+7.13)	78.58 ±3.3 (+6.26)
IDSA+CLUDA	81.88 ±1.7 (+1.24)	54.42 ±3.2 (+0.65)	70.35 ±2.4 (+4.79)	34.50 ±3.3 (+24.06)	75.68 ±2.3 (+3.19)	74.15 ±2.6 (+2.79)
IDSA+ACON	83.77 ±1.6 (+1.70)	62.02 ±2.8 (+3.62)	78.12 ±1.8 (+1.81)	43.55 ±3.1 (+2.83)	**80.80** ±2.3 (+3.84)	**79.05** ±2.3 (+3.54)
IDSA+SASA	84.32 ±1.4 (+1.26)	62.92 ±2.7 (+1.81)	75.78 ±1.9 (+2.55)	41.17 ±2.9 (+1.21)	77.01 ±1.9 (+1.56)	76.23 ±2.3 (+1.12)

Table 2: Cross-subject HAR performance (Acc./F1, %) of source-supervised foundation models and IDSA-integrated UDA methods, averaged over scenarios.

Model	Opportunity	WISDM
MOMENT-small	53.61 / 47.54	59.54 / 28.16
MOMENT-large	54.87 / 49.76	54.83 / 34.75
NormWear	69.54 / 61.13	68.75 / 61.09
Mantis-8M	74.81 / 67.58	76.47 / 70.47
IDSA+Deep CORAL	81.43 / 76.91	75.63 / 63.24
IDSA+CLUDA	84.68 / 84.08	82.14 / 71.67
IDSA+ACON	81.96 / 76.21	79.89 / 64.04
IDSA+SASA	82.96 / 72.23	77.51 / 64.99

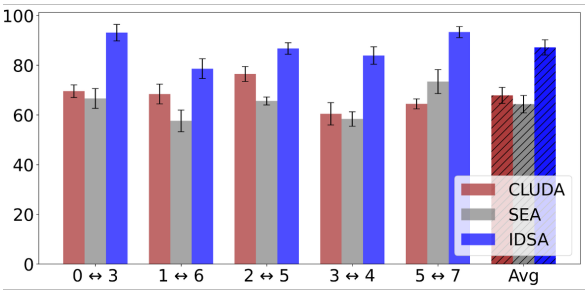


Figure 5: Performance preservation rate under target-side sensor permutations (e.g., $0 \leftrightarrow 3$ swaps the signals of sensors 0 and 3).

on both HAR datasets, demonstrating that sensor-level alignment generalizes across UDA objectives. Specifically, comparing Deep CORAL, SASA, and IDSA+Deep CORAL highlights the advantage of structure-driven strategies, with IDSA further boosting gains via source–target variable correspondence. On Opportunity, IDSA's effect is smaller because diverse multi-location sensors (accelerometers, gyroscopes, magnetometers) yield non-exchangeable channels and cross-unit cost entries in C. Nevertheless, IDSA improves base models in 37 of 48 scenarios and raises average performance for every base model (Table 6), with 10 of the 13 F1-score drops localized to subject 4 target scenarios.

Cross-repetition Model performance on sEMG datasets for cross-repetition adaptation is presented in Table 1, where each domain corresponds to a specific repetition from the same subject. In the sEMG datasets, most domain discrepancies arise due to variations in sensor attachment locations across repetitions within the same subject, making them a natural testbed for sensor-wise domain shift. In this setting, plugging in IDSA yields consistent gains over the corresponding base models across the evaluated sEMG benchmarks. For the Nina5-53 dataset, which exhibits a large gap between accuracy and F1-score due to severe label imbalance, we observe a notable improvement in F1-score when IDSA is added to CLUDA.

6.3 Further Analysis

Comparison with Foundation Models. To assess cross-domain transfer from modern time-series foundation models on wearable MTS, where sensor-wise domain shift can arise, we compare against source-supervised pretrained foundation-model baselines in Table 2. The MOMENT variants show the weakest transfer performance, while NormWear performs substantially better on both HAR benchmarks. Mantis is the strongest foundation-model baseline, particularly on WISDM, but remains below all four IDSA-equipped UDA methods on Opportunity. This suggests that the transferability of pretrained foundation models may vary across downstream domains and models. On WISDM, Mantis outperforms vanilla Deep CORAL, CLUDA, and SASA in accuracy. After adding IDSA, three of the four UDA methods exceed Mantis in accuracy, while IDSA+Deep CORAL

Table 3: Ablation of IDSA components and the LSH approximation on Opportunity (accuracy, %). Cells are shaded by their difference from the baseline (better, worse; saturating at ±8 pp).

Variant	1→2	1→3	1→4	2→1	2→3	2→4	3→1	3→2	3→4	4→1	4→2	4→3	Avg.
CLUDA (baseline)	85.20	86.02	83.65	85.10	77.95	84.45	63.32	76.54	87.67	86.25	85.75	84.47	82.20
+ IDSA, w/o loss	84.91	73.29	76.67	90.25	77.33	74.53	80.80	76.81	81.50	84.81	74.30	78.57	79.48
+ IDSA + LSH approx.	82.68	81.67	80.43	83.95	77.64	81.23	79.37	79.40	82.31	87.97	75.42	83.85	81.33
+ IDSA, exact loss (Ours)	86.03	86.96	84.45	85.96	84.78	81.23	79.08	80.45	82.84	93.69	82.12	88.51	84.68
Deep CORAL (baseline)	81.56	79.50	82.31	84.53	81.67	79.09	81.66	66.20	75.60	81.38	74.58	71.12	78.27
+ IDSA + LSH approx.	84.63	77.01	83.37	87.67	76.09	74.00	82.52	78.21	76.41	87.39	77.65	77.64	80.22
+ IDSA, exact loss (Ours)	86.82	81.06	82.57	89.30	81.37	74.27	83.38	81.01	80.43	83.38	76.82	76.71	81.43

reaches a comparable level. Overall, these results suggest that pretrained representations from foundation models can provide strong transfer in some settings, while target-aware sensor-wise adaptation remains useful.

Sensor Permutation Experiment To further analyze whether IDSA specifically addresses sensor-wise domain shift in controlled synthetic perturbation setting, beyond the naturally occurring cross-subject and cross-repetition settings, we deliberately intensify the correspondence mismatch by swapping the signals of two randomly selected target sensor-channels. Figure 5 shows the performance preservation rate, defined as the ratio between the performance on the permuted target domain and that on the original non-permuted target domain, for the SD-gesture dataset. IDSA with Deep CORAL as the base model is compared with two representative baselines (CLUDA, SEA) across five random sensor permutation cases. The results show that IDSA achieves the highest preservation rate in this diagnostic setting, demonstrating its robustness to sensor permutation, which represents an extreme form of sensor-wise domain shift. While we do not directly quantify ground-truth physical correspondence recovery, the learned transport visualization in Appendix Fig. 10 provides complementary qualitative evidence: under a controlled permutation of target sensors, the learned transport relations change in a manner consistent with the expected swapped-channel correspondence.

Decomposing the gains of IDSA Table 3 separates the contributions of the IDSA module and the proposed loss on Opportunity. Adding the GNN-based IDSA module to CLUDA without the proposed loss lowers the average accuracy from 82.20% to 79.48% (−2.72 pp). The module is therefore not beneficial on its own; its gain depends on the proposed loss, which improves over the w/o loss variant by 5.20 pp on average and yields a net gain of 2.48 pp over the baseline. This decomposition also accounts for the three scenarios in which our method trails the baseline (2→4, 3→4, and 4→2) as shown in Table 6 in the Appendix. In all three, the module alone incurs a large penalty, which the proposed loss narrows the gap but does not fully close it. A large penalty does not by itself imply failure, however: in 1→3, 1→4, and 4→3, the loss recovers enough to surpass the baseline.

Scalability via LSH Approximation Constructing $\mathbf{A}_{st}$ requires all pairwise source–target similarities, which costs $\mathcal{O}(N^2(T+K))$ time and $\mathcal{O}(N^2)$ memory. Time complexity is detailed in Appendix A.6. To reduce this cost, we approximate $\mathbf{A}_{st}$ with locality-sensitive hashing (LSH), computing similarities only for pairs that fall into the same hash bucket. The result of applying LSH approximation is in Table 3. For CLUDA, the LSH approximation outperforms the loss-free IDSA variant in 9 of 12 scenarios but remains 3.35 pp below exact IDSA on average. For Deep CORAL, it outperforms the vanilla baseline in 9 of 12 scenarios and remains 1.21 pp below exact IDSA. The approximation thus trades accuracy for efficiency (e.g., reducing peak memory by 40%).

7 Conclusion

We propose IDSA, a plug-in wearable MTS-UDA module that mitigates *sensor-wise domain shift* via an inter-domain sensor transportation matrix. By explicitly modeling source–target correspondence and regularizing channel redundancy, IDSA aligns target sensor representations with the source domain. Its spatial transportation loss provides a principled, sensor-level discrete optimal-transport objective for cross-domain alignment. Experiments across five benchmarks show that IDSA improves average performance over recent baselines in most evaluated settings, underscoring the importance of sensor-wise alignment for robust wearable domain adaptation.

AI USE STATEMENT

In this work, we used generative AI tools only to edit author-written text for grammar, wording, and readability. We have not used generative AI tools for any of the tasks for which the ICLR 2027 AI Policy for Authors requires disclosure: we did not use them to generate synthetic datasets (the controlled synthetic data in Section 4.2 and Appendix A.5 were generated with our own code), develop theoretical models or conceptual frameworks, formulate mathematical claims or write proofs, propose or refine hypotheses, design or give feedback on the methodology or experiments, implement methods, translate text, clean or reformat datasets, analyze data, or interpret results. Among the tasks with recommended disclosure, we did not use generative AI tools to draft new content, create figures, write or edit code, search or summarize literature, or format references. All AI-suggested edits were reviewed by the authors, who checked that each revision preserved the intended technical meaning and did not alter any equation, numerical result, claim, or citation. We take responsibility for the final content of this work, including any text revised with the aid of generative AI.

ETHICS STATEMENT

This work adheres to the ICLR Code of Ethics. We did not collect new data from human participants; all experiments use four publicly available wearable-sensing datasets (Opportunity, WISDM, NinaPro DB5, and SD-gesture) containing inertial and surface-EMG recordings. We used these datasets as publicly released, with participants identified only by anonymous subject indices; ethical approval and informed consent for the original recordings are described in the respective publications. We note several considerations for downstream use. First, motion and physiological signals can act as biometric identifiers, and IDSA adapts using unlabeled data from each new user or wearing session; a deployed system should collect such target data only with the user's informed consent, preferably processing it on-device. Second, activity and gesture recognition can be misused for monitoring people, such as workers or patients, without their consent; by making recognizers more robust to new users and devices, our method could also make such monitoring easier to deploy, and any deployment should require the consent of the monitored person. Third, although we motivate this work with health monitoring and rehabilitation, we evaluate only benchmark classification performance, which remains limited in some settings; the method has not been validated for clinical or safety-critical use. Finally, the benchmarks include small numbers of participants (e.g., four subjects in Opportunity) and do not cover groups central to these applications, such as older adults or people with motor impairments or limb differences, so the reported gains may not hold equally across user groups.

REPRODUCIBILITY STATEMENT

Our code is available at `https://anonymous.4open.science/r/IDSA_ICLR2027`. Detailed implementation settings, including the feature extractor and hyperparameters, are provided in Appendices A.1 and A.3, with additional ablations and complexity analyses in Appendices A.4 and A.6.

REFERENCES

Toshimitsu Aritake and Hideitsu Hino. Unsupervised domain adaptation for extra features in the target domain using optimal transport. *Neural Computation*, 34(12):2432–2466, 2022.

Oresti Banos, Mate Attila Toth, Miguel Damas, Hector Pomares, and Ignacio Rojas. Dealing with the effects of sensor displacement in wearable activity recognition. *Sensors*, 14(6):9995–10023, 2014.

Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. *Machine learning*, 79(1):151–175, 2010.

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

Nicolas Courty, Rémi Flamary, Devis Tuia, and Alain Rakotomamonjy. Optimal transport for domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(9):1853–1865, 2017.

Yu Du, Wenguang Jin, Wentao Wei, Yu Hu, and Weidong Geng. Surface EMG-based inter-session gesture recognition enhanced by deep domain adaptation. *Sensors*, 17(3):458, 2017. doi: 10.3390/s17030458.

Vasilii Feofanov, Songkang Wen, Shifeng Xie, Simon Roschmann, Marius Alonso, Hongbo Guo, Romain Ilbert, Malik Tiomoko, Quentin Bouniot, Zeynep Akata, Lujia Pan, Jianfeng Zhang, and Ievgen Redko. Mantis: Lightweight foundation model for time series classification. In *Proceedings of the International Conference on Machine Learning*, 2026.

Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *Journal of Machine Learning Research*, 17(1), 2016.

Mononito Goswami, Konrad Szafer, Arjun Choudhry, Yifu Cai, Shuo Li, and Artur Dubrawski. MOMENT: A family of open time-series foundation models. In *Proceedings of the International Conference on Machine Learning*, 2024.

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*, 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 *Proceedings of the International Conference on Machine Learning*, 2023.

Jiaxing Huang, Dayan Guan, Aoran Xiao, Shijian Lu, and Ling Shao. Category contrast for unsupervised domain adaptation in visual tasks. In *Proceedings of the European Conference on Computer Vision*, 2022.

Sozo Inoue, Paula Lago, Tahera Hossain, Tittaya Mairittha, and Nattaya Mairittha. Integrating activity recognition and nursing care records: The system, deployment, and a verification study. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(3), 2019.

Tanguy Kerdoncuff, Rémi Emonet, and Marc Sebban. Metric learning in optimal transport for domain adaptation. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 2020.

Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *Proceedings of the International Conference on Learning Representations*, 2017.

Philip A Knight. The sinkhorn–knopp algorithm: convergence and applications. *SIAM Journal on Matrix Analysis and Applications*, 2008.

Soheil Kolouri, Se Rim Park, Matthew Thorpe, Dejan Slepcev, and Gustavo K. Rohde. Optimal mass transport: Signal processing and machine-learning applications. *IEEE Signal Processing Magazine*, 2017.

Kai Kunze and Paul Lukowicz. Dealing with sensor displacement in motion-based onbody activity recognition systems. In *Proceedings of the 10th International Conference on Ubiquitous Computing*, pp. 20–29, 2008.

Jennifer R Kwapisz, Gary M Weiss, and Samuel A Moore. Activity recognition using cell phone accelerometers. *ACM SIGKDD Explorations Newsletter*, 2011.

Zhiping Lai, Xiaoyang Kang, Hongbo Wang, Weiqi Zhang, Xueze Zhang, Peixian Gong, Lan Niu, and Huijie Huang. STCN-GR: Spatial-temporal convolutional networks for surface-electromyography-based gesture recognition. In *Proceedings of the International Conference on Neural Information Processing*, pp. 27–39, 2021.

Hyeyun Lee, Soyoung Lee, Jaeseong Kim, Heesoo Jung, Kyung Jae Yoon, Srinivas Gandla, Hogun Park, and Sunkook Kim. Stretchable array electromyography sensor with graph neural network for static and dynamic gestures recognition system. *npj Flexible Electronics*, 2023.

Zijian Li, Ruichu Cai, Tom ZJ Fu, Zhifeng Hao, and Kun Zhang. Transferable time-series forecasting under causal conditional shift. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(4):1932–1949, 2024.

Mingyang Liu, Xinyang Chen, Yang Shu, Xiucheng Li, Weili Guan, and Liqiang Nie. Boosting transferability and discriminability for time series domain adaptation. In *Proceedings of the Annual Conference on Neural Information Processing Systems*, 2024.

Qiao Liu and Hui Xue. Adversarial spectral kernel matching for unsupervised time series domain adaptation. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 2021.

Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. Conditional adversarial domain adaptation. In *Proceedings of the Annual Conference on Neural Information Processing Systems*, 2018.

Junxin Lu and Shiliang Sun. CaudiTS: Causal disentangled domain adaptation of multivariate time series. In *Proceedings of the International Conference on Machine Learning*, 2024.

Yunfei Luo, Yuliang Chen, Asif Salekin, and Tauhidur Rahman. Toward foundation model for multivariate wearable sensing of physiological signals. *ACM Transactions on Computing for Healthcare*, 7(3):43:1–43:43, 2026.

Jihye Na, Youngeun Nam, Junhyeok Kang, and Jae-Gil Lee. Mitigating source label dependency in time-series domain adaptation under label shifts. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2025.

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

Stefano Pizzolato, Luca Tagliapietra, Matteo Cognolato, Monica Reggiani, Henning Müller, and Manfredo Atzori. Comparison of six electromyography acquisition setups on hand movement classification tasks. *PloS ONE*, 12(10):e0186132, 2017.

Jinxian Qi, Guozhang Jiang, Gongfa Li, Ying Sun, and Bo Tao. Intelligent human-computer interaction based on surface EMG gesture recognition. *IEEE Access*, 7:61378–61387, 2019.

Ievgen Redko, Nicolas Courty, Rémi Flamary, and Devis Tuia. Optimal transport for multi-source domain adaptation under target shift. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, 2019.

Daniel Roggen, Alberto Calatroni, Mirco Rossi, Thomas Holleczek, Kilian Förster, Gerhard Tröster, Paul Lukowicz, David Bannach, Gerald Pirkl, Alois Ferscha, Jakob Doppler, Clemens Holzmann, Marc Kurz, Gerald Holl, Ricardo Chavarriaga, Hesam Sagha, Hamidreza Bayati, Marco Creatura, and José del R. Millán. Collecting complex activity datasets in highly rich networked sensor environments. In *Proceedings of the International Conference on Networked Sensing Systems*, pp. 233–240, 2010.

Lisa Schrader, Agustín Vargas Toro, Sebastian Konietzny, Stefan Rüping, Barbara Schäpers, Martina Steinböck, Carmen Krewer, Friedemann Müller, Jörg Güttler, and Thomas Bock. Advanced sensing and human activity recognition in early intervention and rehabilitation of elderly people. *Journal of Population Ageing*, 13(2):139–165, 2020.

Shu Shen, Kang Gu, Xin-Rong Chen, Ming Yang, and Ru-Chuan Wang. Movements classification of multi-channel semg based on cnn and stacking ensemble learning. *IEEE Access*, 7:137489–137500, 2019.

Baochen Sun and Kate Saenko. Deep CORAL: Correlation alignment for deep domain adaptation. In *Proceedings of the ECCV 2016 Workshops*, 2016.

Xinwei Tai, Dongmian Zou, and Hongfei Wang. Enhancing node-level graph domain adaptation by alleviating local dependency. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2026.

Rayane Tchantchane, Hao Zhou, Shen Zhang, and Gursel Alici. A review of hand gesture recognition systems based on noninvasive wearable sensors. *Advanced Intelligent Systems*, 5(10): 2300207, 2023.

Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(11), 2008.

Yucheng Wang, Yuecong Xu, Jianfei Yang, Zhenghua Chen, Min Wu, Xiaoli Li, and Lihua Xie. Sensor alignment for multivariate time-series unsupervised domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023.

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 ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2020.

Aaron J. Young, Levi J. Hargrove, and Todd A. Kuiken. The effects of electrode size and orientation on the sensitivity of myoelectric pattern recognition systems to electrode shift. *IEEE Transactions on Biomedical Engineering*, 58(9):2537–2544, 2011.

Aras Yurtman and Billur Barshan. Activity recognition invariant to sensor orientation with wearable motion sensors. *Sensors*, 17(8):1838, 2017.

Xiyuan Zhang, Diyan Teng, Ranak Roy Chowdhury, Shuheng Li, Dezhi Hong, Rajesh Gupta, and Jingbo Shang. UniMTS: Unified pre-training for motion time series. In *Proceedings of the Annual Conference on Neural Information Processing Systems*, 2024.

Chongyang Zhong, Lei Hu, Zihao Zhang, Yongjing Ye, and Shihong Xia. Spatio-temporal gating-adjacency GCN for human motion prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

A Appendix

A.1 Feature Extractor of IDSA

Following prior studies (Ozyurt et al., 2023), we adopt a TCN-based structure for the feature extractor. Therefore, after applying IDSA, the TCN-based feature extractor is applied to extract meaningful representations. The detailed structure of TCN is as follows: for the Opportunity HAR dataset, the feature extractor consists of three layers of TCN structure with a hidden dimension of 64, and five layers with a 32-dimensional hidden dimension for the WISDM dataset.

A.2 Proof of propositions

A.2.1 Definition of discrete 2-Wasserstein Distance

First, we introduce the definition of the discrete version of the Wasserstein distance as follows:

Definition 2 (Discrete 2-Wasserstein distance (Kolouri et al., 2017))**.** *Suppose that when $\mathcal{X}$ and $\mathcal{Y}$ are discrete sets, each having total mass 1 (i.e., each point has mass $1/n$ and $1/m$ respectively). n and m denote the number of data points of $\mathcal{X}$ and $\mathcal{Y}$. Then, using the Euclidean distance, the 2-Wasserstein (or Earth Mover's) distance for the discrete distribution is defined as:*

$$W_2(\mathcal{X}, \mathcal{Y}) = (\min(\sum_{i}^{n}\sum_{j}^{m} \pi_{i,j}\|\mathcal{X}_i - \mathcal{Y}_j\|_2^2))^{\frac{1}{2}} \;\; s.t. \;\; \pi \in \Pi, \tag{13}$$

where Π is $n \times m$ nonnegative coupling matrices whose row sums and column sums are equal to $1/n$ and $1/m$, respectively.

A.2.2 PROOF OF PROPOSITION 1

We extend Proposition 1 as follows:

Proposition 1:Let wearable source MTS data $\mathbf{X}_s$ and target MTS data $\mathbf{X}_t$ be represented as sub-elements in the $\mathcal{X}$ and $\mathcal{Y}$, respectively. For a fixed ground-cost matrix $\mathbf{C}$, optimizing $\mathcal{L}_{\text{st}}$ over the full set of doubly stochastic matrices is equivalent, up to a factor of N, to a discrete 2-Wasserstein problem. In particular, if $\mathbf{A}_{st}$ achieves the global minimum of $\mathcal{L}_{\text{st}}$, it provides an optimal transport plan (up to constant scaling), so $\min \mathcal{L}_{\text{st}} = N \cdot (W_2(\mathbf{X}_s, \mathbf{X}_t))^2$, where $W_2(\mathbf{X}_s, \mathbf{X}_t)$ is the discrete 2-Wasserstein distance between the source and target sensor-channel sets.

Proof. From Equation 6, we introduce the $\mathcal{L}_{\text{st}}$ here as follows, where $(\mathbf{A}_{st})_{i,j}$ represents the element in the i-th row and j-th column of the matrix $\mathbf{A}_{st}$, i.e., the weight with which the j-th target sensor-channel is transported to the i-th source sensor-channel, and $\mathbf{C}_{ij} = \|\mathbf{X}_s(i,\cdot) - \mathbf{X}_t(j,\cdot)\|_2^2$ is the ground cost of Equation 6 under the same index convention (i for source, j for target):

$$\mathcal{L}_{\text{st}} = \sum_{i}^{N} \sum_{j}^{N} (\mathbf{A}_{st})_{i,j} \mathbf{C}_{ij}. \tag{14}$$

Optimizing $\mathcal{L}_{\text{st}}$ leads to the following:

$$\min \mathcal{L}_{\text{st}} = \min \sum_{i}^{N} \sum_{j}^{N} (\mathbf{A}_{st})_{i,j} \|\mathbf{X}_s(i,\cdot) - \mathbf{X}_t(j,\cdot)\|_2^2.$$

Because $\mathbf{A}_{st}$ is doubly stochastic, each row/column sums to 1. Consequently, $\sum_{i,j} (\mathbf{A}_{st})_{i,j} = N$. Define $\widetilde{\pi} = \frac{1}{N} \mathbf{A}_{st}$. Then $\widetilde{\pi}$ has row- and column-sums $1/N$, placing it precisely in the set Π of valid OT couplings in Equation 13, where the N source sensor-channels $\{\mathbf{X}_s(i,\cdot)\}_{i=1}^N$ and the N target sensor-channels $\{\mathbf{X}_t(j,\cdot)\}_{j=1}^N$ play the roles of $\mathcal{X}$ and $\mathcal{Y}$, respectively. Therefore,

$$\sum_{i,j} (\mathbf{A}_{st})_{i,j} \, \|\mathbf{X}_s(i,\cdot) - \mathbf{X}_t(j,\cdot)\|_2^2 = \; N \; \sum_{i,j} \widetilde{\pi}_{i,j} \, \|\mathbf{X}_s(i,\cdot) - \mathbf{X}_t(j,\cdot)\|_2^2.$$

Minimizing the left-hand side with respect to $\mathbf{A}_{st}$ is thus equivalent to minimizing the right-hand side with respect to $\widetilde{\pi} \in \Pi$. The latter is *exactly* the 2-Wasserstein cost up to a factor N. Consequently:

$$\min_{\mathbf{A}_{st}} \; \mathcal{L}_{\text{st}} \; = \; N \; \min_{\widetilde{\pi} \in \Pi} \sum_{i,j} \widetilde{\pi}_{i,j} \; \|\mathbf{X}_s(i,\cdot) - \mathbf{X}_t(j,\cdot)\|_2^2 \; = \; N \; (W_2(\mathbf{X}_s, \mathbf{X}_t))^2.$$

Thus, the statement holds true. □

Empirical analysis of Proposition 1 Building on our theoretical result in Proposition 1, we assess whether the learned inter-domain sensor relation $\mathbf{A}_{st}$ aligns with the ground-truth optimal transport (OT) plan in practice. Specifically, we investigate how closely $\mathbf{A}_{st}$ (trained only under $\mathcal{L}_{\text{st}}$) approximates the exact OT solution on a real sensor dataset.

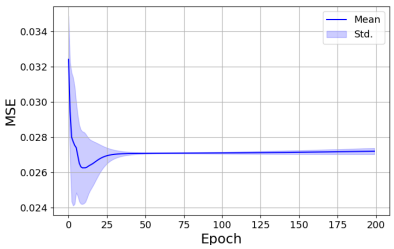


Figure 6: Mean squared error between the learned $\mathbf{A}_{st}$ and the exact OT solution under the specified ground cost π^* on SD-gesture during training (mean and standard deviation over runs).

Setup We employ the SD-gesture sEMG dataset (Lee et al., 2023). To reduce each domain to a representative vector, we compute the mean temporal profile over the batch for each sensor channel. From these profiles, we form a distance matrix $\|\mathbf{X}_s(i,\cdot) - \mathbf{X}_t(j,\cdot)\|_2^2$ for every sensor pair (i, j) and solve a linear assignment problem to obtain a ground-truth OT plan π^*. Meanwhile, we train $\mathbf{A}_{st}$ end-to-end via $\mathcal{L}_{\text{st}}$ alone, as per our method.

Table 4: Hyperparameters to reproduce our sEMG results.

Hyperparameter	**Nina5-18**	**Nina5-53**	**SD-gesture**
lr	0.001	0.001	0.001
batch size	64	128	32
λ_1	1e-1	1e-2	1e-1
λ_2	1e-0	1e-1	1e-1

Table 5: Ablation study of $\mathcal{L}_{st}$ and $\mathcal{L}_{dec}$ on sEMG datasets using Deep CORAL as the base UDA model (accuracy, %). The first row is the vanilla Deep CORAL without IDSA.

$\mathcal{L}_{st}$	$\mathcal{L}_{dec}$	**Nina5-18**	**Nina5-53**	**SD-gesture**
×	×	83.47	75.68	73.02
×	○	82.67	75.83	76.92
○	×	84.98	76.89	79.27
○	○	85.53	78.80	80.15

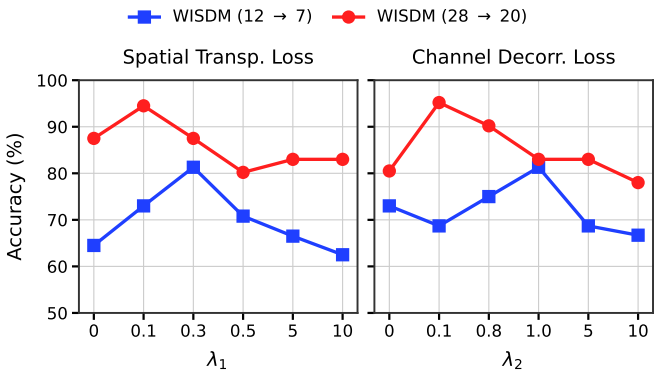


Figure 7: Hyperparameter sensitivity.

Results Figure 6 shows that, during training, the mean squared error between $\mathbf{A}_{st}$ and π^* (both of which have unit row and column sums) decreases rapidly in the early epochs and then plateaus. This indicates that, under $\mathcal{L}_{\text{st}}$, the learned alignment matrix $\mathbf{A}_{st}$ converges toward the optimal transport plan, consistent with Proposition 1.

A.3 Reproducibility and Performance Analysis

A.3.1 Analysis of hyperparameters

To ensure the reproducibility of our model, we list the hyperparameters used in IDSA in Table 4. These parameters were selected based on the validation splits across all datasets. The search space for the learning rate includes [0.001, 0.005, 0.01, 0.05], and for batch size, we consider [32, 64, 128, 256, 512, 1,024, 2048]. For hyperparameters related to the loss functions (λ_1, λ_2), our search space corresponds to [0.01, 0.1, 1, 2, 10]. The embedding dimension of the Temporal layer is experimented with values ranging from 32 to 512, specifically using powers of 2 within this range.

A.4 Extended Experiments

Ablation study Table 5 evaluates the contribution of the $\mathcal{L}_{\text{st}}$ and $\mathcal{L}_{\text{dec}}$. Across all three datasets, the best performance is obtained when both losses are used. The gain from $\mathcal{L}_{st}$ supports the importance of modeling inter-domain sensor relations, while the additional improvement from $\mathcal{L}_{dec}$ suggests that reducing redundant channel couplings provides complementary regularization.

Analysis of hyperparameter sensitivity Figure 7 shows the sensitivity of IDSA to the loss weights λ_1 and λ_2. Performance generally improves up to a moderate value and then slightly de-

Table 6: Overall performance comparison on cross-subject scenarios (source → target). The **bold-faced** score denotes the best result, and the underlined score represents the second-best result. Parentheses denote percentage point gains/losses over base models. For Opportunity HAR, all scenarios are reported, while for WISDM, we follow the same selected scenarios as in Ozyurt et al. (2023).

S → T	Metric (%)	Models													
		Deep CORAL	CDAN	CoDATS	AdvSKM	CLUDA	SEA	RAINCOAT	ACON	SASA	TA4LS	**IDSA + Deep CORAL**	**IDSA + CLUDA**	**IDSA + ACON**	**IDSA + SASA**
Opportunity HAR															
1 → 2	Acc.	81.56	81.56	80.73	83.52	85.20	80.93	82.89	78.98	81.09	81.28	**86.82** (+5.26)	86.03 (+0.83)	85.20 (+6.22)	84.35 (+3.26)
	F1	85.07	84.74	84.10	86.66	87.96	82.11	83.56	83.06	84.65	85.22	**88.69** (+3.62)	88.09 (+0.13)	88.21 (+5.15)	84.78 (+0.13)
1 → 3	Acc.	79.50	78.57	78.57	80.74	86.02	75.31	77.23	75.31	76.87	68.01	81.06 (+1.56)	**86.96** (+0.94)	79.50 (+4.19)	76.40 (-0.47)
	F1	62.11	60.52	60.59	63.26	87.71	55.11	58.02	55.12	51.58	52.71	63.41 (+1.30)	**88.11** (+0.40)	60.16 (+5.04)	58.88 (+7.30)
1 → 4	Acc.	82.31	76.41	**84.72**	83.91	83.65	74.43	75.39	78.13	82.62	78.82	82.57 (+0.26)	84.45 (+0.80)	79.36 (+1.23)	80.16 (-2.46)
	F1	85.78	59.42	87.19	86.40	**88.91**	82.52	58.99	82.62	84.71	60.88	77.11 (-8.67)	88.89 (-0.02)	77.68 (-4.94)	78.38 (-6.33)
2 → 1	Acc.	84.53	82.52	85.67	82.81	85.10	85.62	81.64	87.50	77.22	**93.12**	89.30 (+4.77)	85.96 (+0.86)	89.11 (+1.61)	83.95 (+6.73)
	F1	87.18	62.76	87.88	82.23	87.93	88.75	65.09	88.75	78.45	**94.42**	90.95 (+3.77)	88.38 (+0.45)	91.38 (+2.63)	66.67 (-11.78)
2 → 3	Acc.	81.67	75.16	81.99	80.75	77.95	75.94	77.73	77.50	75.86	80.43	81.37 (-0.30)	**84.78** (+6.83)	80.75 (+3.25)	83.80 (+7.94)
	F1	82.74	58.66	83.42	77.11	81.12	58.47	58.45	60.29	77.38	63.61	79.42 (-3.32)	87.15 (+6.03)	61.15 (+0.86)	**87.50** (+10.12)
2 → 4	Acc.	79.09	82.84	79.36	73.73	**84.45**	76.70	71.48	70.17	82.57	81.77	74.27 (-4.82)	81.23 (-3.22)	73.99 (+3.82)	82.12 (-0.45)
	F1	82.70	76.12	84.48	82.68	**89.32**	53.70	75.45	75.00	74.44	63.02	80.63 (-2.07)	87.45 (-1.87)	68.05 (-6.95)	73.98 (-0.46)
3 → 1	Acc.	81.66	69.91	83.67	77.08	63.32	75.31	70.49	85.54	84.43	**86.53**	83.38 (+1.72)	79.08 (+15.76)	83.67 (-1.87)	83.54 (-0.89)
	F1	62.35	41.64	**84.74**	58.85	51.10	46.60	63.83	53.49	56.20	69.38	63.32 (+0.97)	62.42 (+11.32)	64.20 (+10.71)	60.18 (+3.98)
3 → 2	Acc.	66.20	53.91	62.57	64.25	76.54	67.61	78.52	70.17	67.39	**82.50**	81.01 (+14.81)	80.45 (+3.91)	80.45 (+10.28)	72.07 (+4.68)
	F1	49.46	33.77	50.44	49.64	58.99	44.17	60.28	48.56	50.44	**65.30**	62.23 (+12.77)	61.62 (+2.63)	62.03 (+13.47)	55.68 (+5.24)
3 → 4	Acc.	75.60	64.88	83.11	78.02	**87.67**	66.48	58.20	66.48	80.35	84.18	80.43 (+4.83)	82.84 (-4.83)	81.77 (+15.29)	84.16 (+3.81)
	F1	84.06	72.20	87.32	83.30	**91.39**	72.89	64.00	72.89	72.31	86.53	86.67 (+2.61)	88.09 (-3.30)	86.43 (+13.54)	69.27 (-3.04)
4 → 1	Acc.	81.38	76.79	87.97	86.24	86.25	79.69	87.89	83.13	88.55	91.69	83.38 (+2.00)	**93.69** (+7.44)	85.96 (+2.83)	88.66 (+0.11)
	F1	83.55	71.88	88.99	80.97	88.11	75.26	88.32	81.91	83.11	92.92	83.98 (+0.43)	**94.71** (+6.60)	87.06 (+5.15)	85.51 (+2.40)
4 → 2	Acc.	74.58	68.43	86.31	73.74	85.75	76.14	75.00	77.27	84.04	87.99	76.82 (+2.24)	82.12 (-3.63)	82.96 (+5.69)	**90.20** (+6.16)
	F1	57.42	46.72	88.64	56.43	87.91	78.29	61.56	78.19	63.28	**90.79**	76.62 (+19.20)	84.66 (-3.25)	85.83 (+7.64)	65.59 (+2.31)
4 → 3	Acc.	71.12	72.05	79.19	75.15	84.47	82.81	69.14	80.94	85.14	79.19	76.71 (+5.59)	**88.51** (+4.04)	80.75 (-0.19)	86.06 (+0.92)
	F1	64.11	47.86	60.99	61.49	85.22	81.40	72.19	79.94	64.37	60.64	69.83 (+5.72)	**89.39** (+4.17)	82.29 (+2.35)	80.34 (+15.97)
Avg.	Acc.	78.27	73.59	81.16	78.33	82.20	76.41	75.47	77.59	80.51	82.96	81.43 (+3.16)	**84.68** (+2.48)	81.96 (+4.36)	82.96 (+2.45)
	F1	73.88	59.69	79.07	72.42	82.14	68.27	67.48	71.65	70.08	73.79	76.91 (+3.03)	**84.08** (+1.94)	76.21 (+4.55)	72.23 (+2.15)
WISDM															
2 → 28	Acc.	82.27	71.11	80.92	80.90	84.96	80.00	83.15	75.60	84.94	80.00	**86.67** (+4.40)	**86.67** (+1.71)	82.22 (+6.62)	84.44 (-0.50)
	F1	69.64	44.96	70.93	54.51	80.48	69.35	76.62	74.07	71.81	70.92	78.33 (+8.69)	**83.47** (+2.99)	72.84 (-1.23)	76.67 (+4.86)
7 → 2	Acc.	62.40	75.61	61.06	61.06	78.05	68.29	78.05	**83.33**	71.25	73.17	68.30 (+5.90)	82.93 (+4.88)	82.93 (-0.40)	73.17 (+1.92)
	F1	53.61	51.79	48.04	41.95	59.15	53.61	56.68	60.25	49.98	55.11	55.20 (+1.59)	**73.22** (+14.07)	63.89 (+3.64)	63.46 (+13.48)
7 → 26	Acc.	75.61	73.17	78.05	78.05	73.17	63.41	73.17	81.25	81.08	68.29	78.05 (+2.44)	73.17 (+0.00)	80.49 (-0.76)	**82.05** (+0.97)
	F1	43.19	40.94	47.87	48.51	34.50	33.16	55.31	**59.46**	55.33	40.62	46.76 (+3.57)	39.40 (+4.90)	52.36 (-7.10)	57.42 (+2.09)
12 → 7	Acc.	59.21	70.83	72.10	74.20	79.23	75.00	79.17	81.81	67.30	72.92	83.33 (+24.12)	81.25 (+2.02)	**89.58** (+7.77)	77.08 (+9.78)
	F1	35.73	50.30	65.16	65.61	69.52	50.83	69.21	71.21	50.59	56.08	73.10 (+37.37)	**79.27** (+9.75)	73.14 (+1.93)	50.93 (+0.34)
12 → 19	Acc.	43.33	46.97	63.38	63.98	69.44	54.55	46.97	50.00	73.58	42.42	54.54 (+11.21)	**80.30** (+10.86)	56.06 (+6.06)	75.64 (+2.06)
	F1	23.16	23.04	46.02	46.51	60.70	37.56	31.82	39.22	67.27	25.47	40.52 (+17.36)	**74.93** (+14.23)	39.48 (+0.26)	70.33 (+3.06)
18 → 20	Acc.	38.03	70.73	63.45	39.03	78.03	80.49	76.83	77.35	63.88	**92.68**	68.29 (+30.26)	82.93 (+4.90)	78.05 (+0.70)	75.61 (+11.73)
	F1	29.17	45.83	39.17	40.63	68.29	70.35	64.65	41.78	45.34	**92.44**	43.89 (+14.72)	78.58 (+10.29)	68.81 (+27.03)	53.18 (+7.84)
19 → 2	Acc.	47.35	34.68	39.58	43.45	56.01	41.46	63.41	**81.81**	58.59	58.54	63.41 (+16.06)	65.85 (+9.84)	75.90 (-5.91)	63.41 (+4.82)
	F1	46.21	17.36	34.91	42.43	43.98	52.31	**63.48**	57.85	45.32	55.65	62.34 (+16.13)	47.46 (+3.48)	62.57 (+4.72)	55.63 (+10.31)
28 → 20	Acc.	73.17	75.61	74.10	73.17	80.49	82.93	85.37	82.22	94.76	87.80	**97.56** (+24.39)	87.80 (+7.31)	82.93 (+0.71)	95.12 (+0.36)
	F1	63.33	54.18	53.98	68.25	72.57	67.17	81.14	63.89	84.67	84.87	**97.76** (+34.43)	85.37 (+12.80)	72.36 (+8.47)	90.28 (+5.61)
26 → 2	Acc.	73.77	61.48	72.70	62.01	86.34	53.66	67.07	80.48	76.72	**92.68**	80.50 (+6.73)	90.24 (+3.90)	85.37 (+4.89)	80.24 (+3.52)
	F1	61.78	40.37	59.76	46.30	70.05	37.88	52.48	63.35	69.46	75.40	**82.37** (+20.59)	73.89 (+3.84)	66.53 (+3.18)	72.29 (+2.83)
28 → 2	Acc.	64.92	58.03	71.70	70.69	74.08	65.85	87.80	82.22	67.28	56.10	75.61 (+10.69)	**90.24** (+16.16)	85.37 (+3.15)	68.29 (+1.01)
	F1	49.50	39.98	49.21	48.35	71.06	59.13	72.87	62.88	51.82	52.03	52.16 (+2.66)	**81.11** (+10.05)	68.40 (+5.52)	59.72 (+7.90)
Avg.	Acc.	62.01	63.82	67.70	64.65	75.98	66.56	74.10	77.61	73.94	72.46	75.63 (+13.62)	**82.14** (+6.16)	79.89 (+2.28)	77.51 (+3.57)
	F1	47.53	40.88	51.51	50.31	63.03	53.14	62.43	59.40	59.16	60.86	63.24 (+15.71)	**71.67** (+8.64)	64.04(+4.64)	64.99 (+5.83)

creases, indicating that overly small or large weights can weaken adaptation stability. Overall, values within $[0, 1]$ provide a reasonable balance between performance and robustness across scenarios.

Full results of Figure 4 Table 6 lists the per-scenario accuracy and F1-score behind the averages in Figure 4, covering all 12 source–target pairs of Opportunity HAR and the 10 WISDM pairs selected following Ozyurt et al. (2023). IDSA raises the average accuracy and F1-score of every base model on both datasets, and IDSA+CLUDA attains the best average on both (84.68/84.08 on Opportunity and 82.14/71.67 on WISDM). The gains are largest for Deep CORAL on WISDM (+13.62 accuracy and +15.71 F1 on average, with more than +24 pp in 12→7, 18→20, and 28→20), indicating that a purely statistical alignment objective benefits most from explicit sensor correspondence. Per scenario, IDSA improves accuracy in 37 of 48 Opportunity cases and 35 of 40 WISDM cases (with one tie), and F1-score in 35 of 48 and 38 of 40 cases, respectively. The losses are concentrated: on Opportunity, 10 of the 13 F1 drops occur when subject 4 is the target (1→4, 2→4, 3→4), and on WISDM, three of the four accuracy drops involve ACON in scenarios (7→2, 7→26, 19→2) where ACON is already the strongest non-plug-in baseline, leaving little headroom for further gains.

A.4.1 Visualization of decorrelation Loss

Figure 8 visualizes the channel-wise matrices $\mathbf{D}_s$ and $\mathbf{D}_t$ (Eq. 10) learned by IDSA on Nina5-53. Both matrices are nearly diagonal: the off-diagonal entries are close to zero while the diagonal dominates, indicating that the decorrelation loss suppresses redundant couplings between channels and keeps their representations close to mutually orthogonal. This supports its role in maintaining distinguishable sensor-channel representations for more stable source–target alignment.

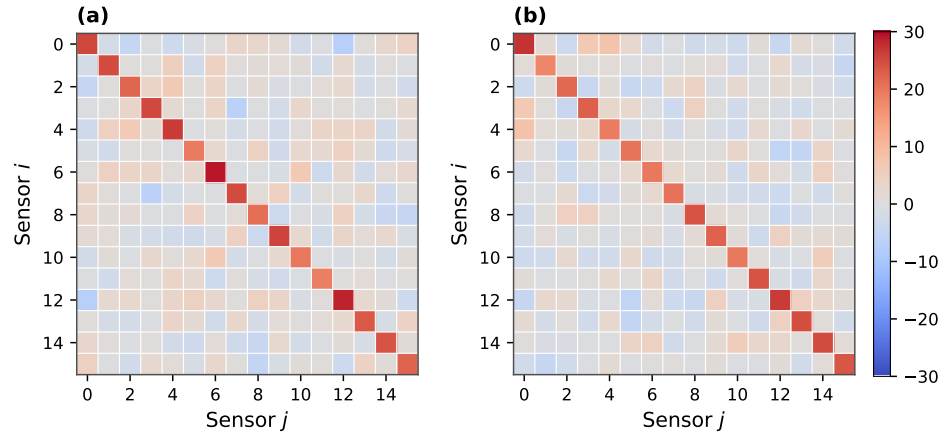


Figure 8: Within-domain sensor relation heatmaps on Nina5-53 dataset. ((a): Source, (b): Target)

Table 7: Comparison of optimal-transport-based UDA methods.

Method	OT granularity	Modality	Task
MLOT (Kerdoncuff et al., 2020)	Instance-level	Generic vector/image	UDA
OTDA (Courty et al., 2017)	Instance-level	Generic vector/image	UDA
DAEVS (Aritake & Hino, 2022)	Instance-level	Generic vector/sensor (target-only extra features)	UDA with dimensional mismatch
RAINCOAT (He et al., 2023)	Instance-level	Time-series	UDA with and without label shift
IDSA (Ours)	Sensor-level	Multivariate time-series	UDA with sensor-wise domain shift

A.4.2 Discussion with Optimal Transport-based Methods

There is a line of works adopting optimal transport (OT) for domain adaptation (Kerdoncuff et al., 2020; Courty et al., 2017; Aritake & Hino, 2022). Table 7 summarizes the conceptual comparison across prior OT-based domain adaptation methods and IDSA. OTDA (Courty et al., 2017) and MLOT (Kerdoncuff et al., 2020) align sample distributions in a feature space whose coordinates are assumed to have stable semantics across domains. DAEVS (Aritake & Hino, 2022) addresses a heterogeneous setting where the target has extra features, applying OT on the shared part while modeling the extra part with pseudo-labeling. RAINCOAT (He et al., 2023), which we already include as a baseline, aligns latent time and frequency features via Sinkhorn divergence and corrects possible label shift. In contrast, our IDSA treats each sensor channel as an entity and performs OT at the **sensor level** by learning a sensor channel-wise transport map to resolve sensor-wise domain shift, which is illustrated in our motivation. Thus, our approach differs from instance-level OT methods and instead addresses a distinct and commonly encountered problem in wearable MTS.

We additionally implement the MLOT (Kerdoncuff et al., 2020) and OTDA (Courty et al., 2017) models for a more thorough performance comparison of OT-based DA methods, while excluding DAEVS (Aritake & Hino, 2022), whose setting is different from ours, requiring an additional target dimension. The results are presented in Table 8. From the results, we can infer that RAINCOAT, which models OT specifically for time-series data, performs better than the other two traditional OT-based UDA methods. Our methodology, which models the transportation map for sensor-level optimal transport, still demonstrates the highest performance even when compared to newly adopted traditional OT methods.

A.4.3 Spatial transportation visualization

We visualize the inter-domain transport map on SD-gesture under two settings: source–original target and source–permuted target, where target sensors 3 and 6 are exchanged. Figure 10(a) corresponds to the source–original target scenario, while Figure 10(b) corresponds to the source–permuted target scenario in which target sensors 3 and 6 are exchanged. In bipartite graphs, red nodes denote source-domain sensors and yellow nodes denote target-domain sensors, and edge thickness indicates the learned transport weight. For each scenario, we show the full graph and a focused view highlighting edges incident to the two target sensors. The highlighted relations for sensors 3

Table 8: Performance comparison with OT baselines on HAR datasets (Opportunity, WISDM).

S → T	Metric (%)	Models				
		MLOT	OTDA	RAINCOAT	IDSA + Deep CORAL	IDSA + CLUDA
Opportunity						
1 → 2	Acc.	74.30	70.39	82.89	86.82	86.03
	F1	76.66	66.25	83.56	88.69	88.09
1 → 3	Acc.	69.25	69.88	77.23	81.06	86.96
	F1	42.32	42.58	58.02	63.41	88.11
1 → 4	Acc.	72.39	75.64	75.39	82.57	84.45
	F1	79.96	68.85	58.99	77.11	88.89
2 → 1	Acc.	83.38	75.93	81.64	89.30	85.96
	F1	85.92	72.24	65.09	90.95	88.38
2 → 3	Acc.	70.19	69.88	77.73	81.37	84.78
	F1	65.05	42.58	58.45	79.42	87.15
2 → 4	Acc.	73.46	70.95	71.48	74.27	81.23
	F1	81.81	67.79	75.45	80.63	87.45
3 → 1	Acc.	67.62	67.91	70.49	83.38	79.08
	F1	40.07	40.37	63.83	63.32	62.42
3 → 2	Acc.	65.08	62.57	78.52	81.01	80.45
	F1	40.32	42.20	60.28	62.23	61.62
3 → 4	Acc.	56.84	55.50	58.20	80.43	82.84
	F1	41.90	50.97	64.00	86.67	88.09
4 → 1	Acc.	70.20	72.49	87.89	83.38	93.69
	F1	58.69	58.62	88.32	83.98	94.71
4 → 2	Acc.	68.99	65.92	75.00	76.82	82.12
	F1	50.01	40.99	61.56	76.62	84.66
4 → 3	Acc.	69.88	69.88	69.14	76.71	88.51
	F1	59.15	59.11	72.19	69.83	89.39
Avg.	Acc.	70.13	68.91	75.47	81.43	84.68
	F1	60.16	54.38	67.48	76.91	84.08
WISDM						
2 → 28	Acc.	82.22	80.00	83.15	86.67	86.67
	F1	59.60	81.52	76.62	78.33	83.47
7 → 2	Acc.	65.85	65.85	78.05	68.30	82.93
	F1	52.01	50.84	56.68	55.20	73.22
7 → 26	Acc.	73.17	73.73	73.17	78.05	73.17
	F1	38.96	36.27	55.31	46.76	39.40
12 → 7	Acc.	64.58	64.58	79.17	83.33	81.25
	F1	55.14	71.22	69.21	73.10	79.27
12 → 19	Acc.	54.55	71.21	46.97	54.54	80.30
	F1	41.38	67.81	31.82	40.52	74.93
18 → 20	Acc.	41.46	36.59	76.83	68.29	82.93
	F1	32.84	36.53	64.65	43.89	78.58
19 → 2	Acc.	58.54	51.22	63.41	63.41	65.85
	F1	51.94	51.60	63.48	62.34	47.46
28 → 20	Acc.	70.73	75.61	85.37	97.56	87.80
	F1	65.51	54.83	81.14	97.76	85.37
26 → 2	Acc.	63.31	53.66	67.07	80.50	90.24
	F1	48.93	35.63	52.48	82.37	73.89
28 → 2	Acc.	60.98	63.41	87.80	75.61	90.24
	F1	48.93	52.76	72.87	52.16	81.11
Avg.	Acc.	63.54	63.59	74.10	75.63	82.14
	F1	49.52	53.90	62.43	63.24	71.67

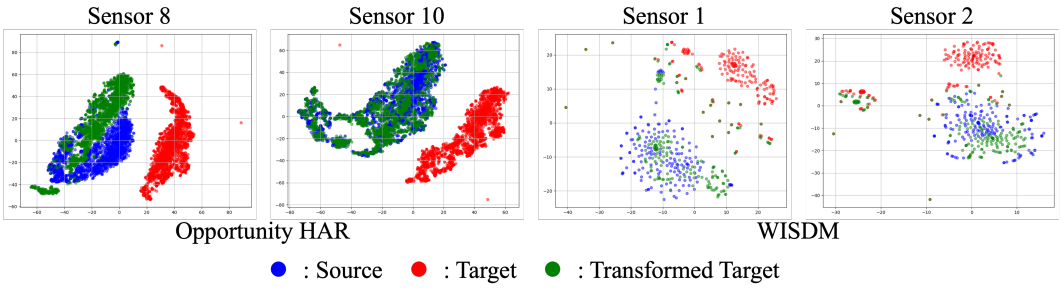
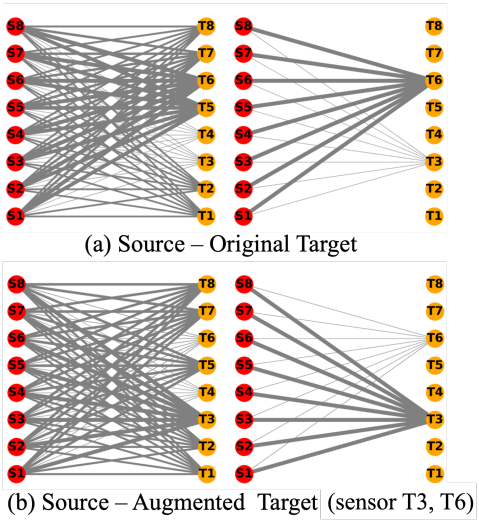


Figure 9: t-SNE visualization of selected sensor-channel representations for the source, target, and transformed target domains.

and 6 differ across the two scenarios in a way that is consistent with the permutation, while most other source–target relations remain largely stable. This suggests that IDSA can adaptively capture inter-domain sensor correspondence under sensor-wise domain shift.

Sensor-aware visualization We further visualize source, target, and transformed target sensor-channel representations using t-SNE. Figure 9 presents t-SNE (Van der Maaten & Hinton, 2008) results for sensors 8 and 10 from Opportunity HAR and sensors 1 and 2 from WISDM. While the original source and target representations are clearly separated, the transformed target representations move closer to the source after applying the learned transport matrix. These results suggest that inter-domain transformation reduces sensor-level discrepancy and makes target data more compatible with the source-trained model.



(a) Source – Original Target

(b) Source – Augmented Target (sensor T3, T6)

Figure 10: Visualization of the learned transportation map as a bipartite graph over source and target sensor-channels.

A.5 Detail of Section 4.2

A.5.1 Empirical analysis on controlled synthetic datasets

In real data, $\mathbf{M}_d$ is latent and entangled with other sources of variation. Therefore, we generate synthetic signals from Equation 2 with prescribed $\mathbf{M}_d$, allowing us to vary the mixing matrix while holding the shared latent $\mathbf{U}$ fixed and thereby probe the observations from the previous section. We conduct a stress test with representative *structure-aware* UDA (feature alignment with a spatial encoder; Deep CORAL (Sun & Saenko, 2016)) and *structure-driven* UDA (explicit structure alignment; SASA (Cai et al., 2021)) baselines under progressively stronger sensor-wise domain shift by modeling three types of shifts as follows:

- **Cross-subject sensor orientation shift:** We group 18 channels into 6 sensors with 3 axes each and apply a block-diagonal rotation, with severity controlled by the rotation angle θ (degrees).
- **Cross-repetition sensor placement shift:** We arrange 8 channels on a circular electrode ring and model rotation by a circulant channel mixing operator parameterized by the rotation degree θ.
- **Variable permutation:** For the 8 arranged channels, we apply permutations with increasing severity, measured by the number of permuted channels k.

For each synthetic shift, we define a source domain a using the identity transform (no correspondence shift), and create target domains $b, c, d, \ldots$ by applying shifts in the channel space. All domains share the same class-conditional temporal generator.

Figure 2(a) summarizes target accuracy as a function of shift severity. Within each of the three plots, the correspondence shift is progressively strengthened from left to right according to the UDA scenarios. Both Deep CORAL and SASA achieve strong performance under mild correspondence shift, but their accuracy degrades as the shift becomes more severe. This controlled trend is consistent with our analysis: when source–target sensor correspondence becomes increasingly unstable, methods relying primarily on feature alignment or invariant within-domain structure may fail to reduce $\epsilon_T(h)$.

A.5.2 Empirical analysis on real-world MTS

This section shows that *sensor-wise domain shift* is implicitly present in real-world wearable data and that applying a simple correspondence-based remedy can improve adaptation performance. We randomly select 10 sensor-channels from two subjects in the Opportunity HAR dataset (Roggen et al., 2010). To analyze within-domain and cross-domain variable relations, we adopt cosine similarity between sensor-channel signals as a pairwise relation measure. A simple *heuristic* $\mathbf{A}_{s\leftarrow t}\mathbf{X}_t$ is constructed as follows: for each target sensor channel j, we replace the signals of the j-th sensor channel with a weighted sum of the top-3 source sensor channels that show the highest similarity to it. We denote the target and transformed target data as $\mathbf{X}_t$ and $\mathbf{X}_t^{\text{trans}}$, respectively. We then evaluate a UDA baseline Deep CORAL (Sun & Saenko, 2016) with a GNN as the encoder under four settings: Baseline $(\mathbf{I}, \mathbf{X}_t)$, Within-Only $(\Sigma_t, \mathbf{X}_t)$, Cross-Only $(\mathbf{I}, \mathbf{X}_t^{\text{trans}})$, and Cross+Within $(\Sigma_t, \mathbf{X}_t^{\text{trans}})$.

Figure 2(b) shows that applying the correspondence-based target transformation improves target accuracy in this diagnostic setup (Cross-only $>$ Baseline; Cross+Within $>$ Within-only). These results support the need to estimate source–target variable correspondence in a data-dependent manner for each domain pair, rather than relying solely on within-domain sensor relations.

A.6 Time Complexity of $\mathcal{L}_{\text{ST}}$

Constructing the exact inter-domain transport requires pairwise computations over N source and N target sensor-channels. Since each channel is represented by a length-T temporal profile, computing both the sensor-similarity matrix and the ground-cost matrix requires $\mathcal{O}(N^2T)$ time. Sinkhorn normalization further requires $\mathcal{O}(KN^2)$ time for K iterations, resulting in an overall complexity of $\mathcal{O}(N^2(T+K))$ and $\mathcal{O}(N^2)$ memory for the dense pairwise matrices. In our benchmarks, N ranges from 3 to 113 for HAR and from 8 to 16 for sEMG, keeping the exact computation practical for the evaluated settings. In cases where the number of sensors is large, we additionally evaluate the LSH-based approximation described in Section 6.3, which restricts pairwise computations to candidate sensor pairs, and efficiency–accuracy trade-off is reported in Table 3.