Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Unsupervised Cross-Domain Adaptation for Wearable-based Human Activity Recognition

Indrajeet Ghosh

vkw4ze@virginia.edu

University of Virginia

Garvit Chugh

Indian Institute of Technology Jodhpur

Abu Zaher Md Faridee

Amazon (United States)

Nirmalya Roy

University of Maryland, Baltimore County

Research Article

Keywords: Unsupervised Domain Adaptation, Conditional Alignment, Temporal Ensembling, Action Recognition, Consistency Regularization, Data Augmentation

Posted Date: March 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8622641/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Unsupervised Cross-Domain Adaptation for Wearable-based Human Activity Recognition

Indrajeet Ghosh[1*], Garvit Chugh[2], Abu Zaher Md Faridee[3], Nirmalya Roy[4]

[1*]Department of Systems and Information Engineering, University of Virginia, USA.
[2]Department of Computer Science and Engineering, Indian Institute of Technology Jodhpur, India.
[3]Amazon, USA.
[4]Department of Information System, University of Maryland Baltimore County, USA.

*Corresponding author(s). E-mail(s): vkw4ze@virginia.edu;
Contributing authors: chugh.2@iitj.ac.in; abufari@amazon.com;
nroy@umbc.edu;

Abstract

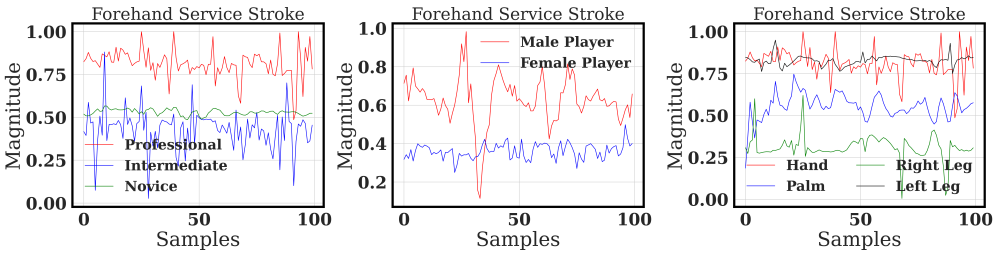
Recent advancements in deep learning-based wearable human action recognition (wHAR) have improved the capture and classification of complex motions. Still, adoption remains limited due to the lack of expert annotations and domain discrepancies from user variations. Limited annotations hinder the model's ability to generalize to out-of-distribution samples. While data augmentation can improve generalizability, unsupervised augmentation techniques must be applied carefully to avoid introducing noise. Unsupervised domain adaptation (UDA) addresses domain discrepancies by aligning conditional distributions with labeled target samples, but vanilla pseudo-labeling can lead to error propagation. To address these challenges, we propose *CoHAR*, a novel joint optimization architecture comprised of three functions: (i) consistency regularizer between augmented samples to improve model classification generalizability, (ii) temporal ensemble for robust pseudo-label generation and (iii) conditional distribution alignment to improve domain generalizability. The temporal ensemble works by aggregating predictions from past epochs to smooth out noisy pseudo-label predictions, which are then used in the conditional distribution alignment module to minimize kernel-based class-wise conditional maximum mean discrepancy (***k***CMMD) between the source and target feature space to learn a domain invariant embedding. The

consistency-regularized augmentations ensure that multiple augmentations of the same sample share the same labels; this results in (a) strong generalization with limited source domain samples and (b) consistent pseudo-label generation in target samples. The novel integration of these three modules in *CoHAR* results in a range of $\approx$ ***4-12%*** average macro-F1 score improvement over six state-of-the-art UDA methods in five benchmark wHAR datasets.

Keywords: Unsupervised Domain Adaptation, Conditional Alignment, Temporal Ensembling, Action Recognition, Consistency Regularization, Data Augmentation

1 Introduction

Over the last decade, wearable-based human activity recognition (wHAR) has ushered in numerous applications in healthcare [1, 2], sports analytics [3, 4], fitness tracking [5, 6], cognitive studies [7, 8] and many more [9–11]. This growth is propelled by the integrated inertial, gyroscopic, and magnetometric sensors in smartphones and smartwatches, which offer an unprecedented ability to capture human motion data [12]. The neural processing units in state-of-the-art wearables enhance real-time deep neural network inference for various assessments. However, traditional models rely on supervised learning, requiring extensive labeled data, which is a barrier to developing scalable and robust activity recognition models.



(a) Different expert levels (b) Gender variations (c) Cross-position variations

Fig. 1: Example of cross-user (expertise and gender) and cross-positions variations in BAR [4] dataset.

Activities of daily living (ADLs) datasets capture common activities like *running, walking, sitting, etc.* However, for micro-complex activities, the patterns are often unique to individual users, leading to significant cross-user variations [13] due to individual differences in body characteristics and skill levels [4]. Such disparities are particularly pronounced in actions requiring learning and coordination, where user proficiency can change over time. For instance, as badminton players develop skills like smashing or volleying, the labels from their earlier-less skilled performances may

no longer accurately represent their current abilities. This is evident when comparing data patterns of experienced players, who exhibit more consistent and distinctive patterns, against those of novices or intermediate players.

We empirically illustrate via accelerometer data from badminton players of varying skill levels and genders in the BAR dataset [4], showcasing distinct patterns between professional, intermediate, novice players (Figure 1a), genders (Figure 1b) and cross-positions (Figure 1c). We observe significant fluctuations in magnitudes among professional players, ranging from **[0.52 - 0.98]**, in contrast to intermediate **[0.18 - 0.88]** and novice **[0.48 - 0.54]** players. Gender-based differences are also noted, with male players showing magnitudes between **[0.12 - 0.97]**, while female players exhibit a narrower range of **[0.33 - 0.48]**. Similarly, such magnitude variations can be observed during cross-position analysis, where signals from four sensor positions on the body are plotted. For instance, the hand and left leg exhibit higher magnitudes **[0.6–1.0]** and **[0.6–0.9]**, respectively, while the palm and right leg show lower magnitudes **[0.3–0.7]** and **[0.1–0.5]**, respectively, illustrating substantial variability across sensor locations. These findings underscore the challenge of capturing and classifying diverse movement patterns across users, skill levels, and sensor locations, which can degrade the generalizability and robustness of the framework.

In this context, unsupervised domain adaptation (UDA) [14, 15] presents itself as a viable solution, especially when annotations are scarce or absent and there is a need to manage domain discrepancies arising from out-of-distribution samples. UDA utilizes labeled data from users in the training set (source domain) and unlabeled data from new domains (target domain). It aligns feature distributions across source (e.g. female, expert player) and target domains (e.g. male, novice player) to learn a domain-invariant feature space. Recent advancements in UDA models [16–21], specifically for cross-user wHAR, have been promising. Existing UDA research [17, 22, 23] mainly focuses on aligning marginal feature distributions, with few models aligning conditional distributions by assuming access to labeled target data [24]. An alternative is to use pseudo-labeling techniques, but its self-labeling nature often propagates the initial classification errors [25], complicating conditional distribution alignment between domains.

To address these challenges, we introduce a novel joint optimization framework where each of the loss components is carefully chosen to complement the others. Our strategy begins with training a deep learning classifier on labeled data from the source users. We then try to adapt this model with the target domain unlabeled data by learning a domain-invariant embedding by minimizing the class-wise kernel-based conditional maximum mean discrepancy (kCMMD) distance. Since the target domain samples do not provide labeled data, we rely on pseudo-labels from the target domain classifier. We employ a fast temporal ensemble approach to prevent training collapse due to error propagation from incorrect pseudo-labels, where the aggregate function smoothes out noisy pseudo-labels. To improve model generalizability, we employ a label-preserving data augmentation approach that incorporates prior class information for both source and target data streams; it is accompanied by consistency training in the form of minimizing KL divergence between predictions from the real and augmentation samples, leading to an interesting synergy with the previously mentioned

modules. Consistency training ensures that multiple augmentations of the same sample always receive the same label by the classifier, allowing the source model to become robust against small noises in data with limited labeled samples and cover some uncertain regions in the decision space (i.e. out of the initial source distribution). Similarly, it ensures that the pseudo-labels generated over the target domain's unlabeled samples remain the same regardless of the augmentation, and it harmonizes with the temporal ensemble to provide high-quality pseudo-labels for the conditional domain alignment module to minimize the domain gap and improve target domain classification without any labeled samples. This synergy allows *CoHAR*to achieve considerable performance benefits over *six* state-of-the-art unsupervised domain adaptation baselines in *five* activity recognition datasets that cover simple activities of daily living and complex sports/gym actions.

The key contributions are summarized as follows:

- We propose a novel unsupervised cross-domain adapting framework (*CoHAR*) to address two critical challenges in the wHAR area: (i) *cross-user (gender, expertise) and cross-position variations* and (ii) *limited labeled data* by learning source-specific representations from labeled data and leveraging the learned representation to unlabeled target user data.
- We develop a novel joint optimization architecture with (i) *kernel-based class-wise conditional maximum mean discrepancy* (kCMMD) to align conditional distributions between the source and target domain features, (ii) *temporal ensemble-based pseudo-labeling* technique to provide robust labels for the target domain and (iii) *consistency regularizer between label-preserving augmentations* to improve model generalizability, robustness against noise and improve the quality of target domain pseudo-labels. We detail the synergistic interaction of these complementary components with thorough ablation studies.
- We compare the performance (macro-F1) of *CoHAR*against *six* state-of-the-art (SOTA) algorithms [26–31] in *five* public benchmark datasets [4, 32–35] that incorporate various levels of complexity in the label space (simple ADLs vs micro-complex sport/gym activities) as well as diversity in dexterity levels (e.g. novice vs expert player), gender and cross-position. We show that *CoHAR*constantly outperforms the baselines by a range of **4-12%** in 5 independent trials per dataset.

2 Related Work

In this section, we discuss the relevant literature to *CoHAR* on wearable-based human action recognition (wHAR), unsupervised domain adaptation in the wHAR and consistency regularization area with an emphasis on handling cross-user/-position variations and limited labeled data, and try to differentiate *CoHAR*from the state-of-the-art approaches.

2.1 Wearable-based Human Action Recognition

Wearable sensors have seen widespread use in various sports, including tennis [36], cricket [37], basketball [38], and gym activities [39]. Although there have been significant advances in wearable-based recognition of daily human activities [40], scaling these models to more niche areas where annotations are scarce has been an open challenge, especially for activities that require expert execution like sports and gym activities. In [41], the authors propose a sophisticated action recognition framework that leverages wearable sensors for badminton sports. This framework adopts a transfer learning strategy to achieve user-independent generalization across different users for badminton sports. However, this study did not address real-world issues such as limited label availability and cross-user variations. In [42], propose a contrastive predictive coding (CPC) framework to human activity recognition, which captures and learns the representation of the long-term temporal structure of the dataset. CPC module comprises three modules; pre-training is utilized to learn the representation from the unlabeled data, followed by a fine-tuning module, which performs activity recognition tasks on the learned features from the pre-training module and a softmax to classify activities. Similarly, research from [43] and [44] explores the impact of sensor number and placement on gym activity recognition, but they did not address the challenge of minimizing inter-class variations. While several other studies have concentrated on the ADLs and iADLs human actions [45, 46], our work uniquely extends tackle both cross-user variations and limited label data for wHAR domain.

2.2 Wearable-based Unsupervised Domain Adaptation (UDA)

UDA has become a valuable method for overcoming the challenge of limited labeled data as it capitalizes on the abundance of available unlabeled data to enhance the model's generalizability and robustness. One of the few works incorporating UDA to tackle limited label data is [26], where the authors introduced a novel sample differentiation technique, utilizing a parameterized network to identify whether a sample is from the source or target domain and additionally involves assigning weights to pseudo labels for target samples based on the confidence level of the domain classifier. In [47], the authors investigate domain distribution shifts by incorporating statistical alignment and uncertainty-aware coupling mechanisms. Specifically, uncertainty estimation via entropy minimization is used to guide the coupling between source and target representations, thereby enabling more reliable pseudo-labeling under domain shift. Furthermore, in [16], conducted a comprehensive comparison between three adaptation strategies for wHAR: feature matching, confusion maximization, and data augmentation across various human activity recognition datasets. They provide evidence that data augmentation and feature-matching approaches enhance robustness and adaptability considerably. However, recent SOTA studies have focused on addressing minimal label information issues via minimizing conditional feature distribution. In contrast, this work aims to tackle the limited label data by minimizing conditional discrepancies.

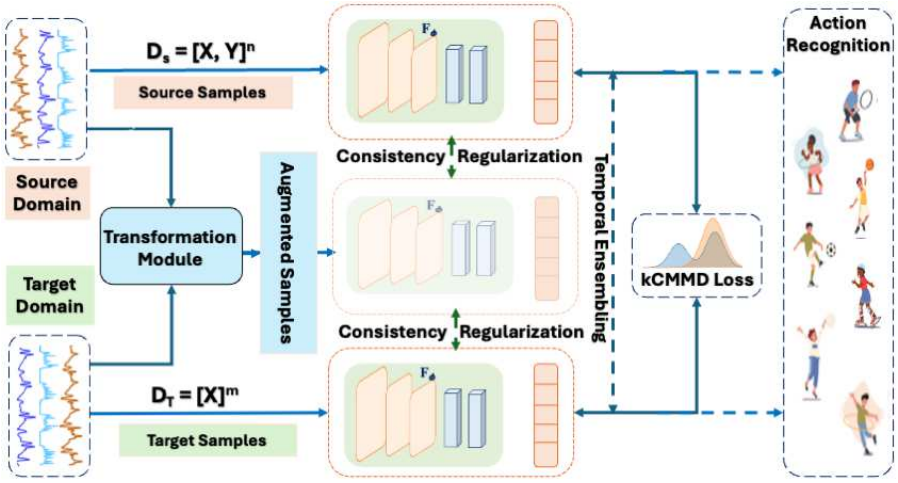


Fig. 2: High level overview of the *CoHAR*architecture comprising four key modules: (i) Supervised training of source-specific features representation (ii) Extraction pseudo labels for unlabeled target domain (iii) Incorporating kernel-wise class-wise kCMMD loss to minimize the conditional shift between the domains and (iv) Consistency regularizer for data augmentation.

2.3 Consistency Regularization

Consistency regularization has recently emerged as a key paradigm for improving model robustness by encouraging invariant predictions under diverse data perturbations, including augmented inputs [48], noisy training conditions [49], and bias-aware consistency constraints [50]. The core principle of consistency regularization is to maintain consistent predictions between real and augmented samples. This approach promotes smoothness and low-density separation, crucial for semi-supervised learning environments, aiding in minimizing consistency loss [51]. Studies [52, 53] indicate that combining consistency regularization with unsupervised data augmentation in semi-supervised learning impedes performance degradation and bolsters model generalizability and robustness. Additionally, [54] shows that encouraging invariance by reducing distances between feature distributions of differently augmented data can enhance classification performance. Motivated by this, we integrate a consistent regularization approach to tackle user variations. However, to design real-world settings, we incorporate the three key principles of unsupervised data augmentation to generate diverse samples for the wHAR domain [55], and to the best of our knowledge, our work is among the first few that have integrated unsupervised geometric-based data augmentation with consistency regularization in the wHAR domain.

3 Methodology

In this section, we delve into the intricacies of our proposed UDA framework, *CoHAR*, tailored for recognizing ADL actions, encompassing gym- and sports-related actions

as well. We begin by outlining the problem formulation and then highlight and discuss the functional and learning components in the *CoHAR*overall architecture module.

3.1 Problem Formulation

The challenge of cross-user variations and limited expert-labeled datasets highly impede the capability of wHAR. Our approach leverages labeled data from experienced users (source domain), represented as $D_s = [x_i, y_i]_{i=1}^{n_s}$, and unlabeled data from new, unseen users (target domain), denoted as $D_t = [x_j]_{j=1}^{n_t}$. With n_s and n_t indicating the number of instances in the source and target domains, respectively, both domains share the same label space, defined as $y_s^i, y_t^i \in \{1, 2, \ldots, c\}$, where 'c' represents the number of activity classes. The underlying premise for employing UDA in this context is to establish a framework that can accurately identify the activities of new users by exploiting the labeled data from the source domain alongside the unlabeled data from the target domain. This presupposes that both the source and target users utilize identical sensors and sensor placements, performing the same set of activities, despite the inherent distribution shift between the domains, i.e., $P(x_s) \neq P(x_t)$.

Figure 2 highlights the pictorial overview of the proposed UDA framework, which is intricately designed around four pivotal components to address these challenges effectively: (i) learning source-specific feature representations to capture the distinctive characteristics of the trained user data accurately; (ii) generating high-quality pseudo labels through temporal ensembling, enhancing label accuracy by leveraging temporal data consistency; (iii) achieving conditional alignment between the source and target domain features using kCMMD, a sophisticated technique for fine-tuning the adaptation process; and (iv) implementing data augmentation to bolster the model's generalizability and robustness against variations and noise inherent in cross-user (gender and expertise) data. This integrated approach aims to minimize domain discrepancies, facilitating the model's ability to generalize across new users in the wHAR domain.

3.2 Supervised Learning Source-specific Feature Learning

We explore the supervised training process designed to capture the unique features of the source domain data. Our approach involves a specifically tailored supervised training architecture. This architecture consists of three convolutional layers, each with an increasing number of filters and progressively smaller kernel sizes. Each convolutional layer is coupled with batch normalization and max-pooling layers to minimize internal covariate shifts and enhance learning of the high and low-level features, respectively. Followed by two fully connected layers, and a softmax layer for output. The training objective is to minimize the cross-entropy supervised loss by employing labeled training data, as defined in Equation (1). This process aims to accurately predict the class distribution, denoted as $P(\hat{Y}|X)$. We organize the dataset for both the source and target domains into separate batches for efficient processing.

$$\mathcal{L}_{SL} = -\frac{1}{n}\sum_{i=1}^{n} P(y_{s_i}|x_{s_i}) \log P(\hat{y}_{s_i}|x_{s_i}) \tag{1}$$

3.3 Extracting Pseudo Labels using Temporal Ensembling

Adapting our supervised training pipeline from a well-labeled source domain to a target domain devoid of labeled data presents a significant challenge. This is further complicated by the inherent variability and complexity of activities, especially when dealing with data from synchronized sensors on multiple wearable devices. To effectively generate high-quality pseudo labels in this context, we implement *temporal ensembling* [56]. This method is designed to maximize the concurrent extraction of mutual information from both the source and target domains, thereby enhancing pseudo-label generation by iteratively aggregating model predictions.

At each training iteration t, the model parameters θ_{t+1} are updated via a gradient step while an exponential moving average of predictions from previous iterations is maintained using a decay factor α shown in Eq. (3) and (2), respectively. *A higher value of α places greater emphasis on historical ensemble predictions, smoothing out noisy fluctuations, whereas a lower α incorporates recent changes more rapidly, enabling quicker adaptation to evolving model outputs.* Let $\mathbf{P}_t$ be the predictions at iteration t, and $\mathbf{E}_t$ denote the ensemble of historical predictions and the temporal ensembling updates $\mathbf{E}_t$ as follows:

$$\mathbf{E}_t = \alpha\, \mathbf{E}_{t-1} \;+\; (1-\alpha)\, \mathbf{P}_t \tag{2}$$

$$\theta_{t+1} = \theta_t - \eta \cdot \nabla_\theta \mathcal{L} \tag{3}$$

where $\mathcal{L}$ is the loss function, θ represents the model parameters, η is the learning rate, and ∇_θ denotes the gradient update. By aggregating predictions $\mathbf{E}_t$, temporal ensembling smooths the overall label estimation for unlabeled target data, improving pseudo-label quality and reducing label noise. It also addresses class imbalance [57] by enabling more stable learning in the presence of minority classes, facilitating a balanced, domain-invariant feature representation. For further details, Alg. 1 outlines the pseudocode for temporal ensembling. We determine the optimal setting for α via a grid-search approach, ensuring that the chosen value is fine-tuned for optimal performance across all benchmark datasets.

3.4 UDA via kernel-based class-wise conditional mean maximum discrepancy (kCMMD)

Our focus is on aligning conditional distributions using our proposed kernel-based class-wise conditional mean maximum discrepancy (kCMMD) approach, particularly within the Reproducing Kernel Hilbert Space (RKHS) [58]. In RKHS, conditional distributions are mapped to a high-dimensional feature space. Here, kCMMD calculates the mean discrepancy of these embeddings to minimize domain shifts, aiming to reduce the discrepancy between the source and target domains. The kCMMD approach diminishes the conditional probability gap between the source domain ($P^s(Z|Y)$) and the target domain ($P^t(Z|Y)$), thus establishing a robust conditional domain-invariant feature space. The generalized kCMMD loss is defined in Eq. 4, where C represents the number of classes, and n_c and m_c denote the number of samples from class c in the source and target domains, respectively, with x_i^s, x_j^s, x_i^t, and x_j^t representing feature embeddings.

$$\mathcal{L}_{\text{kCMMD}} = \sum_{c=1}^{C} \left[\frac{1}{n_c(n_c-1)} \sum_{\substack{i,j=1 \\ i\neq j}}^{n_c} K(\phi)(x_i^s, x_j^s) + \frac{1}{m_c(m_c-1)} \right. \\ \left. \sum_{\substack{i,j=1 \\ i\neq j}}^{m_c} K(\phi)(x_i^t, x_j^t) - \frac{2}{n_c m_c} \sum_{i=1}^{n_c} \sum_{j=1}^{m_c} K(\phi)(x_i^s, x_j^t) \right] \tag{4}$$

where: $(\phi(\cdot))$ represents the feature map to a RKHS, (n_c^s) and (n_c^t) are the numbers of samples in the source and target domains for each class (c), respectively, (x_i^s) and (x_j^t) are the samples from the source and target domains corresponding to labels $(y_i^s = c)$ and $(y_j^t = c)$ and (C) is the total number of classes considered across both the domains.

Considering our problem scenario, where the source domain D_s is represented by a feature set $x_s \in \mathbb{R}^{n_s \times d}$ with corresponding labels y_s, and a target domain D_t represented by $x_t \in \mathbb{R}^{n_t \times d}$ with pseudo-labels $y_{\bar{t}}$, the kCMMD loss focuses on minimizing distribution differences for each class $c \in C$, where C is the set of classes. kCMMD utilizes the Radial Basis Function (RBF) kernel, defined as $K(\phi)(x, x') = \exp(-\gamma|x - x'|^2)$, where γ is the kernel bandwidth parameter. For each class c from source (X_s) and target (X_t) domains, we compute the following kernel matrices (K_M), quantifying intra- and inter-domain discrepancies:

$$\text{Source-Source } [K_M^{ss}] = K(\phi)(x, x') \otimes (X_s^{(c)}, X_s^{(c)}, \gamma) + \lambda I \tag{5}$$

$$\text{Target-Target } [K_M^{tt}] = K(\phi)(x, x') \otimes (X_t^{(c)}, X_t^{(c)}, \gamma) + \lambda I \tag{6}$$

$$\text{Source-Target } [K_M^{st}] = K(\phi)(x, x') \otimes (X_s^{(c)}, X_t^{(c)}, \gamma) \tag{7}$$

Here, λI denotes the regularization constant added to the diagonals of the source-source and target-target kernel matrices to prevent issues like ill-conditioning and to control overfitting [58, 59]. This regularization ensures that the learned features are adaptable and generalizable across users. Kernel methods are instrumental in capturing non-linear features, and mapping data into higher-dimensional spaces where non-linear relationships can be linearly separated. The kCMMD loss is calculated as the mean of discrepancies across all classes using a function δ, which computes the mean discrepancy for each class. The kCMMD discrepancy function for the kernel matrices K_M^{ss}, K_M^{tt}, and K_M^{st} is defined in Equation (8):

$$\mathcal{L}_{kCMMD} = (\bar{K_M^{ss}} + \bar{K_M^{tt}} - 2 \cdot \bar{K_M^{st}}) \tag{8}$$

Equation 8 measures the overall conditional discrepancy between the source and target domain distributions. K_M^{ss} and K_M^{tt} represent the mean intra-domain similarities for the source and target domains, respectively, while K_M^{st} quantifies the mean inter-domain similarity. The equation calculates domain discrepancy by adding the mean similarities within each domain and subtracting twice the mean cross-domain similarity, i.e. ensuring a bidirectional symmetrical step to seek to minimize differences between the source and target domains without bias. When the source and target

domain distributions align closely, the values of K_M^{ss} and K_M^{tt} will approximate K_M^{st}, resulting in a lower domain discrepancy value for effective cross-user adaptation task.

3.5 Data Augmentation Module

In this part of our framework, we address the challenge of user variability, such as differences in activity execution, proficiency levels, gender, and individual traits. We integrate the core idea fo unsupervised data augmentation approach [52] to create a robust model against these cross-user variations. The augmentation module is developed to ensure three key properties: (i) *augmented samples must be valid and realistic compared to real samples*, (ii) *should exhibit diversity*, and (iii) *should introduce specific inductive biases*.

Our objective is to tackle various invariances commonly found in temporal wearable data [60] across users. We recognize that sudden temporal shifts in the data can significantly impair the performance of our model. To counteract this, work is dedicated to developing a model that remains invariant to such transformations and augmentations. To achieve this, we implement two types of geometric-based augmentations—jitter and rotation. These methods are well-established for effectively capturing the real-world variations encountered in wHAR tasks [55, 61].

- **Jitter (σ)**: Introduces high-frequency fluctuations in the data, distinct from actual motion, thereby adding stochastic noise. This variation challenges the interpretation of finer movements or subtle orientations detected by wearable sensors.
- **Rotation (r_θ)**: Involves angular displacement or orientation change, quantifying the degree and axis of rotational movement. This is vital for understanding dynamic orientation changes during ADLs/iADLs actions.

Let G represent the geometric-based augmentation applied to the source and target domain data, i.e. x_s and x_t. The augmented samples are then defined as $G = [r_\theta, \sigma]$ and $x_s^{'}, x_t^{'} = (G \cdot (x_s), G \cdot (x_t))$. Our key aim is maintaining consistency between the model's predictions on original and augmented data, ensuring invariance to user-specific variations. We employ a consistency loss for the source and target domains, as defined in Equations (9), to enforce model stability across augmented or transformed source and target data samples, respectively. Here, $f_\theta(x_s)$, $f_\theta(x_s')$, $f_\theta(x_t)$ and $f_\theta(x_t')$ represent the model predictions for the real and augmented samples from source and target domains, respectively. This approach aligns with the principles of the Lipschitz continuity theorem [62], ensuring that the model's output changes are proportionate to the sample variations. This aspect is critical for enhancing the model's robustness and generalization in handling diverse user-specific data variations.

$$\begin{aligned}\mathcal{L}_{\mathrm{C}} = \mathbb{E}x_s \sim D_s, x_s' \sim t(x_s)\left[D_{KL}(f_\theta(x_s)||f_\theta(x_s'))\right] + \\ \mathbb{E}x_t \sim D_t, x_t' \sim t(x_t)\left[D_{KL}(f_\theta(x_t)||f_\theta(x_t'))\right]\end{aligned} \tag{9}$$

Algorithm 1 Pseudo Code for ***CoHAR*** Training

Require: Source domain $D_s = \{(x_i, y_i)\}_{i=1}^{n_s}$, Target domain $D_t = \{x_j\}_{j=1}^{n_t}$, ensemble predictions E, counts C, new predictions P, indices I, smoothing factor α
Ensure: Trained model parameters θ, pseudo-labels $\tilde{y}_t$

Initialize model parameters θ
Initialize ensemble predictions E and counts C
for each training epoch **do**
  for each batch B_s in D_s **do**
    Perform supervised training on B_s to update θ
    Augment B_s to obtain B_s': $B_s' \leftarrow G(B_s)$
    Perform unsupervised training on B_s' using pseudo-labels $\tilde{y}_s$
  end for
  for each batch B_t in D_t **do**
    Predict on B_t using current θ to obtain P
    for each index i in batch B_t **do**
      $E[i] \leftarrow \alpha \cdot E[i] + (1 - \alpha) \cdot P[i]$
      $C[i] \leftarrow C[i] + 1$
    end for
    Compute averaged ensemble prediction:
    $A \leftarrow \frac{E}{C}$.unsqueeze(1)
    Generate pseudo-labels: $\tilde{y}_t \leftarrow \arg\max(A, \text{dim} = 1)$
    Augment B_t to obtain B_t': $B_t' \leftarrow G(B_t)$
  end for
  Compute $\mathcal{L}_{SL}$ loss ***Supervised Loss***
  Compute $\mathcal{L}_{kCMMD}$ loss ***Domain Adaptation***
  Compute $\mathcal{L}_C$ loss ***Consistency Regularization***
end for
return θ, $\tilde{y}_t$

3.6 *CoHAR* Training Procedure

The overall aim is to obtain a set of optimized model parameters trained on labeled source data that work well in the target domain and generate high-quality pseudo-labels for the target domain. To generate high-quality pseudo labels, we utilize the ensembling approach due to the ability to smoothen noisy pseudo labels. Additionally, we adopted a conditional probability alignment approach, which provides a robust domain-invariant feature embedding. We highlight the overall joint optimization objective function for *CoHAR*framework enumerated in Equation (10) and Algorithm 1 showcase the pseudo-code of the overall *CoHAR*pipeline, respectively.

$$\min_{F_\theta} \mathcal{L}_{overall} = \underbrace{\mathcal{L}_{SL} + \beta_0 \mathcal{L}_{kCMMD}}_{\text{Domain Adaptation}} + \underbrace{\beta_1 \mathcal{L}_C}_{\text{Consistency Regularizer}} \tag{10}$$

In the overall formulation equation, θ encompasses all the parameters within the deep network. The hyperparameters β_0 and β_1 are set to a numerical value of 1, facilitating the fine-tuning of the overall objective function. This approach ensures a balanced integration of conditional domain adaptation and consistency regularization within CoHAR.

4 Experiments

We evaluate *CoHAR* via extensive experiments on the problem of limited labeled settings for ADLs action recognition by investigating the following research questions (RQs):

- **RQ1 (Performance Analysis of *CoHAR*):** How well is *CoHAR*performing on publically available datasets?
- **RQ2 (Robustness Analysis of *CoHAR*):** How can we adapt *CoHAR*for accurate recognition across cross-position and cross-gender in wHAR?
- **RQ3 (Comparative Analysis of *CoHAR*):** Is *CoHAR*minimizing conditional data distribution discrepancies?
- **RQ4 (Compatibility Analysis of *CoHAR*):** Is *CoHAR*optimized for high-quality pseudo-label generation?
- **RQ5 (Effectiveness Analysis of *CoHAR*):** (a) Impact of augmentation on CoHAR generalizability and (b) λ regularizer parameter on conditional domain alignment and also *CoHAR* convergence highlighting optimal performance?

4.1 Setup

4.1.1 Datasets

We selected five publicly available datasets to demonstrate our framework's scalability and adaptability across a wide range of activities, from basic Activities of Daily Living (ADLs) to complex activities with significant temporal shifts and erratic movements. (1) **Badminton Activity Recognition (BAR) Dataset** [4] aiming to assess micro-complex actions within badminton based on badminton strokes from 11 subjects in both controlled (indoor facility, selected for expertise) and uncontrolled (open playground, randomly selected) environments. It includes 30 iterations of 12 strokes. (2) **Daily and Sports Activity (DSADS) Dataset** [32] comprising data from eight individuals (balanced gender distribution) engaging in 19 diverse activities. The dataset features 9 Degrees of Freedom (DOF) sensor data from accelerometers, magnetometers, and gyroscopes, positioned on five body locations and recorded at 25 Hz. (3) **Physical Activity Monitoring (PAMAP2) Dataset** [33] containing data from nine subjects performing 18 physical activities, with data captured at 100 Hz from three Colibri sensors placed on the wrist, ankle, and chest. The dataset comprises activities ranging from walking, playing soccer, cycling, etc. (4) **Workout Activities (MM-DOS) Dataset** [34] involving 50 participants across expertise levels (expert, intermediate, beginner) performing four gym workouts (squats, lunges, push-ups, shoulder presses). The focus is on IMU data from nine body locations recorded at 50 Hz. (5) **MPSC-AR Dataset** [35] comprise of **5** subjects performing **6** ADL/I-ADL activities (*sitting*, *walking*, *washing utensils*, *folding laundry*, *brushing teeth*, and *writing*) and only accelerometer data collected at 32 Hz is used.

4.1.2 Baselines

We evaluate and contrast the performance of *CoHAR*with SOTA UDA algorithms including SWL-Adapt[26], ContrasGAN [27], AdvSVM [29], HoMM [30], DUA [31] and CoTMix [28] (1) (4) **SWL-Adapt** leverages meta-optimization and pseudo-labeling to dynamically weight samples, enabling adaptive domain discrimination, whereas (2) **ContrasGAN** uses bi-directional GANs for domain-invariant feature learning and reduce the domain discrepancies with contrastive loss (3) **CoTMix** innovates with a temporal mixup for domain augmentation, emphasizing cross-domain similarity through contrastive learning learning domain-invariant feature space (4) **AdvSKM** comprise of hybrid spectral kernel network can precisely learn the non-stationary and non-monotonic statistics of the data distribution and befits for domain matching (5) **HoMM** leverages higher-order moment matching in RKHS to reduces the domain discrepancy via source domain loss minimization, higher-order domain alignment and discriminative clustering with pseudo-labeled target samples. (6) **DUA** uses a novel mechanism for continuous domain adaptation by dynamically adjusting batch normalization statistics (mean and variance) using a minimal unlabeled dataset.

4.2 Implementation Details

4.2.1 Data Pre-Processing

This study uses raw signals from accelerometers, gyroscopes, and magnetometers in a body-worn IMU sensor network as input features. To address noise issues like motion artifacts, we applied a median filter for preprocessing and normalized 48 features using a min-max scaler for the *BAR* dataset, which includes data from three axes of each sensor type. Given the clipping of signals beyond ($\pm 2g$) due to the low-noise accelerometer's limited range, we opted for wide/high-range accelerometer signals to capture rapid movements better. For feature extraction, we implemented a sliding window technique with varying overlap rates and sampling rates across the *BAR*, *PAMAP2*, *DSADS*, *MPSC-AR* and *MMDOS* datasets to remove artifacts and extract temporal patterns effectively. Labeling was done using majority voting within each window segment. We also used a *grid-search* method to fine-tune the optimal window sizes and overlap percentages, exploring a range from 5% to 95% overlap and window durations from 0.1 to 1 second, enabling us to select the most suitable parameters for capturing complex data patterns across different datasets.

4.2.2 Network Architecture and Training

Our experiments were executed on a Linux server equipped with an Intel i7-6850K CPU, 4x NVIDIA GeForce GTX 1080Ti GPUs, and 64 GB RAM. For data preprocessing and deep learning tasks, we utilized Python, specifically focusing on PyTorch libraries for the deep learning components. Due to the imbalance of class labels in most datasets, we use the macro-F1 score as the primary evaluation metric. The datasets were divided into training, validation, and testing sets with a 60-20-20% split. Best hyperparameter optimization values, highlighted in (Table 4), were conducted using the validation set, ensuring that neither the validation nor the testing sets were

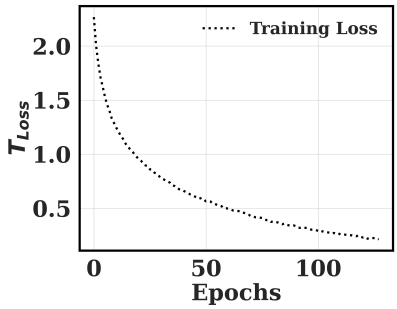
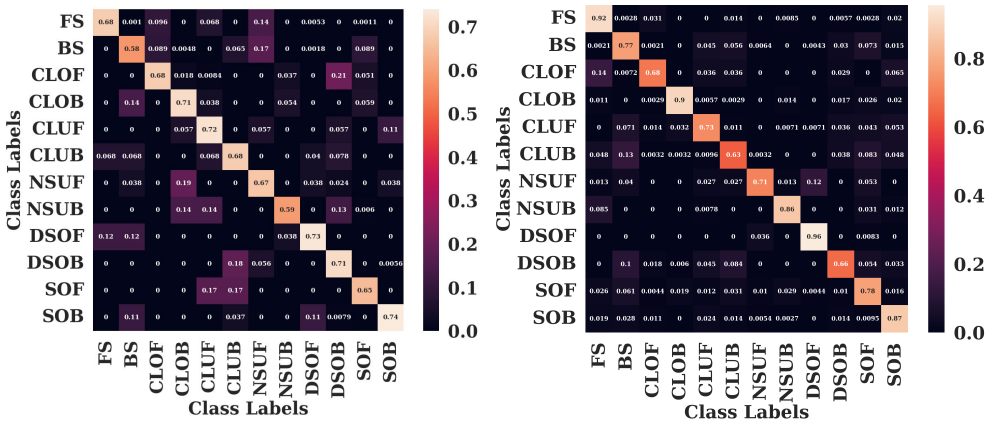


Fig. 3: Total Training Loss Convergence

Hyper-parameter	Values
No. of maximum convolution layers	1,2,3
No. of filters in convolution layers	32, 128, 64
Convolution filter dimension	5x1,5x1,5x1
No. of maximum fully connected layers	2
No. of neurons in fully connected layers	8, 4
Batch size	64
Dropout rate	0.3
Optimizer	Adam
Learning Rate	0.001 - 0.0003
Max number of epochs	128
λ value	0.18 - 0.45
α value	0.55 - 0.75
σ value	0.01 - 0.10
r_θ	$\left[-\frac{\pi}{4}, \frac{\pi}{4}\right]$

Fig. 4: List of *CoHAR's* Hyperparameters

involved in the training process. Notably, *CoHAR*was fine-tuned and optimized solely through hyperparameter tuning on the validation set, yielding the best recognition performance and improving model generalization. Figure 3 shows the overall training loss (T_{Loss} **represents** $\mathcal{L}_{SL} + \beta_0 \mathcal{L}_{kCMMD} + \beta_1 \mathcal{L}_C$) of the *CoHAR*convergence on the source and unlabeled target training data. *CoHAR*achieves stable performance within 128 epochs, learning domain-invariant features from source and target concurrently while minimizing inter-domain discrepancy.



(a) CoTMix results (b) *CoHAR*results

Fig. 5: Confusion matrices on the BAR dataset and labels are the acronyms of various Badminton Strokes: **F**orehand **S**ervice, **B**ackhand **S**ervice, **C**lear **L**ob **O**verhead **F**orehand, **C**lear **L**ob **O**verhead **B**ackhand, **C**lear **L**ob **U**nderarm **F**orehand, **C**lear **L**ob **U**nderarm **B**ackhand, **N**et **S**hot **U**nderarm **F**orehand, **N**et **S**hot **U**nderarm **B**ackhand, **D**rop **S**hot **O**verhead **F**orehand, **D**rop **S**hot **O**verhead **B**ackhand, **S**mash **O**verhead **F**orehand and **S**mash **O**verhead **B**ackhand.

Table 1: Comparison of CoHAR's Macro F1 Score (± SD) and Average Improvement Margin vs. SWL-Adapt [26], ContrasGAN [27], CoDATS, CoTMix [28], AdvSKM [29], HoMM [30], and DUA [31] across BAR, DSADS, PAMAP2, and MMDOS datasets. The best and second-best results are shown in bold and underline, respectively.

Algorithms	**DSADS**	**PAMAP2**	**MMDOS**	**BAR**	**MPSC-AR**
SWL-Adapt [26]	0.70 ± (0.0188)	0.73 ± (0.0174)	0.72 ± (0.0181)	0.68 ± (0.0219)	0.74 ± (0.0119)
ContrasGAN [27]	0.64 ± (0.0213)	0.72 ± (0.0160)	0.78 ± (0.0159)	0.71 ± (0.0192)	0.68 ± (0.0142)
CoTMix [28]	0.61 ± (0.0252)	0.66 ± (0.0248)	0.68 ± (0.0227)	0.73 ± (0.0205)	0.75 ± (0.0081)
AdvSKM [29]	0.71 ± (0.0132)	0.70 ± (0.0213)	0.69 ± (0.0193)	0.70 ± (0.0125)	0.69 ± (0.0149)
HoMM [30]	0.68 ± (0.0265)	0.67 ± (0.0229)	0.70 ± (0.0162)	0.65 ± (0.0217)	0.72 ± (0.0132)
DUA [31]	0.73 ± (0.0105)	0.72 ± (0.0095)	0.76 ± (0.0122)	0.69 ± (0.0155)	0.67 ± (0.0218)
***CoHAR* (Our)**	**0.77 ± (0.0239)**	**0.78 ± (0.0211)**	**0.84 ± (0.0256)**	**0.82 ± (0.0198)**	**0.81 ± (0.0069)**
Average Improvement Margin	**0.048 ± (0.015)**	**0.054 ± (0.011)**	**0.061 ± (0.013)**	**0.092 ± (0.019)**	**0.063 ± (0.009)**

Table 2: Cross-gender adaptation male (source) → female (target).

Datasets	CoTMix [28]	SWL-Adapt [26]	ContrasGAN [27]	HoMM [30]	DUA [31]	AdvSKM [29]	***CoHAR*(Our)**
BAR [4]	0.61 ± (0.0235)	0.57 ± (0.0201)	0.61 ± (0.0161)	0.63 ± (0.0202)	0.67 ± (0.0122)	0.68 ± (0.0341)	**0.73 ± (0.0163)**
DSADS [32]	0.67 ± (0.0212)	0.61 ± (0.0433)	0.69 ± (0.0361)	0.62 ± (0.0281)	0.64 ± (0.0361)	0.66 ± (0.0288)	**0.70 ± (0.0.254)**

Table 3: Cross-gender adaptation female (source) → male (target).

Datasets	CoTMix [28]	SWL-Adapt [26]	ContrasGAN [27]	HoMM [30]	DUA [31]	AdvSKM [29]	***CoHAR*(Our)**
BAR [4]	0.73 ± (0.026)	0.68 ± (0.029)	0.70 ± (0.031)	0.63 ± (0.0281)	0.71 ± (0.0111)	0.65 ± (0.0198)	**0.77 ± (0.0148)**
DSADS [32]	0.71 ± (0.034)	0.66 ± (0.023)	0.65 ± (0.014)	0.62 ± (0.0299)	0.67 ± (0.0119)	0.70 ± (0.0207)	**0.76 ± (0.0189)**

4.3 Results and Discussion

Here, we highlight and enumerate the key findings and discuss in-depth the results of our experiments and research questions, analyzing the effectiveness and robustness of *CoHAR*.

4.3.1 Performance Analysis of *CoHAR* (RQ1)

In this analysis, we contrast and compare against the state-of-the-art (SOTA) UDA algorithms, demonstrating that *CoHAR* outperforms the existing SOTA UDA algorithms by a significant margin, achieving a **4-12%** improvement on average across five public datasets. Our analysis distinguished the top-performing and second-best results, revealing a notable trend: SWL-Adapt frequently excels beyond other leading UDA algorithms across benchmark datasets, which might be attributed to utilizing a meta-optimized loss function, which effectively assigns sample weights, enhances domain alignment, and significantly aids in the cross-user adaptation, as shown in Table 1.

Notably, CoHAR performance stands out across all datasets, encompassing activities ranging from complex actions like gym and sports-related activities to ADLs and iADLs. The framework's strength lies in its use of temporal ensembling for generating reliable, high-confidence pseudo-labels and integrating kernel-based class-wise conditional mean maximum discrepancy (kCMMD) with temporal ensembling strengthens *CoHAR*'s domain adaptation capabilities. Additionally, we notice a

notable improvement in overall performance when utilizing consistency regularization to tackle augmented samples' inherent variations in real-world settings. Further analyses supporting these findings are presented in section 4.3.3.

Interestingly, in the class-wise representation via confusion matrix, **CoTMix** demonstrates exceptional performance compared to state-of-the-art (SOTA) UDA algorithms, particularly in handling the inherent complexities and data variations of the BAR dataset. This effectiveness is attributed to its robust encoder and class-aware contrastive training, combined with a temporal mixup strategy that introduces controlled variability, enhancing its capability in complex action recognition tasks, as illustrated in figure 5a. This strategy allows CoTMix to better capture domain-agnostic features across complex patterns in both source and target domains. Similarly, ***CoHAR*** also significantly outperforms the aforementioned SOTA UDA methods and exhibits a comparable trend, as shown in figure 5b, demonstrating its ability to effectively handle data variations and inter-class heterogeneities.

Lastly, we conduct a hypothesis test to validate the statistical significance of the observed performance improvements, following non-parametric statistical analysis [63]. Overall, the results confirm that *CoHAR* achieves robust and statistically supported performance gains across diverse domain-shift scenarios, with a significant improvement over state-of-the-art baselines ($p = 0.02471 < 0.05$).

Qualitative ablation result: We validate *CoHAR* on the BAR dataset's target user using a t-SNE plot that illustrates class embeddings in a two-dimensional space, as shown in Figure 6. This visualization reveals that inter-class activities cluster closely, while intra-class variations diverge in a structured manner, highlighting *CoHAR*'s efficacy in distinctively embedding complex actions, such as badminton in our case. This observation underscores the model's capability to learn distinct complex action patterns and its proficiency in separating intra-class boundaries, thereby demonstrating robust class separation and feature representation under cross-user adaptation in the wHAR domain. Through this analysis, we reaffirm that *CoHAR* effectively distinguishes complex actions (sports- and gym-based activities), while also accurately recognizing ADL patterns such as sitting and walking, where variations are comparatively lower than those observed in complex motor activities. Additionally, to complement this qualitative analysis, we compute the Proxy A-distance [64] to quantify domain discrepancy between the source and target users. The Proxy A-distance decreases from 1.79±0.03 before adaptation to 0.84±0.02 after applying *CoHAR*, indicating a substantial and stable reduction in cross-user domain discrepancy. Together,

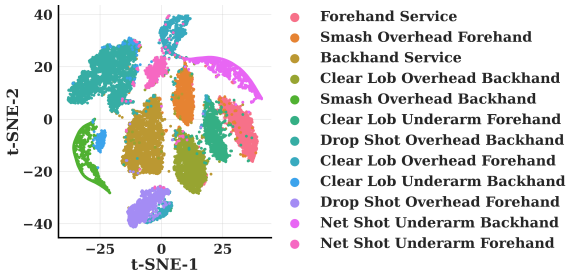
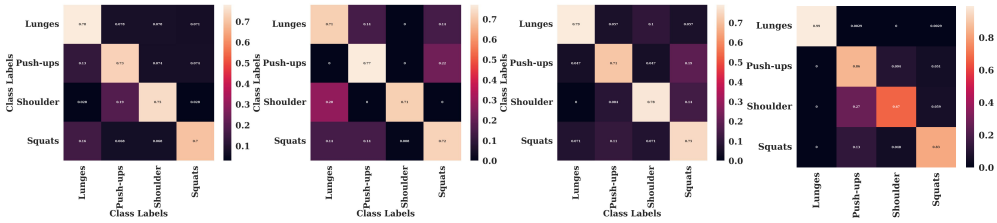


Fig. 6: Feature Embedding on Target User using *CoHAR* on the BAR dataset.



(a) ContrasGAN [27] (b) CoTMix [28] (c) SWL-Adapt [26] (d) *CoHAR*(Our)

Fig. 7: Confusion matrices for cross-dexterity levels (expert (source) → beginner (target)) comparing SWL-Adapt and *CoHAR*algorithms on MMDOS dataset.

Table 4: Cross-body adaptation results where **upper body** corresponds to sensor locations such as hand and palm (BAR), left and right upper arms, left and right wrists (MMDOS), right and left arms (DSADS), and wrist (PAMAP2); and **lower body** corresponds to sensor locations such as right and left legs (BAR), left and right calves, left and right thighs (MMDOS), right and left legs (DSADS), and ankle (PAMAP2).

Algorithms	BAR	MMDOS	DSADS	PAMAP2	MPSC-AR
CoTMix [28]	0.71 ± 0.0202	0.78 ± 0.0212	0.73 ± 0.0112	0.74 ± 0.0123	0.70 ± (0.0192)
ContrasGAN [27]	0.70 ± 0.0214	0.72 ± 0.0194	0.70 ± 0.0198	0.71 ± 0.0208	0.68 ± (0.0229)
DUA [31]	0.77 ± 0.0289	0.75 ± 0.0263	0.72 ± 0.0271	0.72 ± 0.0098	0.74 ± (0.0126)
SWL-Adapt [26]	0.79 ± 0.0095	0.80 ± 0.0106	0.75 ± 0.0209	0.76 ± 0.0213	0.79 ± (0.0091)
HoMM [30]	0.74 ± 0.0217	0.77 ± 0.0105	0.73 ± 0.0378	0.70 ± 0.0301	0.73 ± (0.0231)
AdvSKM [29]	0.76 ± 0.0111	0.74 ± 0.0118	0.71 ± 0.0099	0.77 ± 0.0207	0.76 ± (0.0110)
***CoHAR* (Our)**	**0.80 ± 0.0312**	**0.81 ± 0.0121**	**0.76 ± 0.0234**	**0.79 ± 0.0119**	**0.83 ± 0.0102**
Average Improvement	**≈ 0.0102**	**≈ 0.0147**	**≈ 0.0129**	**≈ 0.0202**	**≈ 0.0409**

(a) Upper (source) to Lower (target)

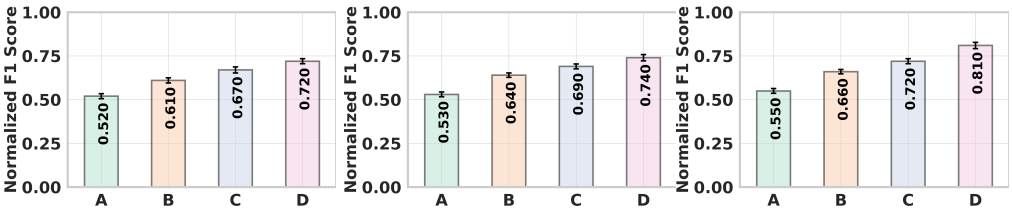
Algorithms	BAR	MMDOS	DSADS	PAMAP2	MPSC-AR
CoTMix [28]	0.71 ± 0.0295	0.68 ± 0.0301	0.73 ± 0.0091	0.74 ± 0.0103	0.72 ± (0.0118)
ContrasGAN [27]	0.73 ± 0.0376	0.71 ± 0.0101	0.72 ± 0.0286	0.70 ± 0.0098	0.70 ± (0.0181)
DUA [31]	0.75 ± 0.0076	0.77 ± 0.0117	0.69 ± 0.0213	0.69 ± 0.0201	0.72 ± (0.0223)
SWL-Adapt [26]	0.76 ± 0.0197	0.73 ± 0.0101	0.71 ± 0.0209	0.73 ± 0.0187	0.77 ± (0.0144)
HoMM [30]	0.72 ± 0.0091	0.74 ± 0.0112	0.70 ± 0.0209	0.75 ± 0.0108	0.73 ± (0.0118)
AdvSKM [29]	0.70 ± 0.0569	0.71 ± 0.0198	0.72 ± 0.0271	0.72 ± 0.0203	0.76 ± (0.0122)
***CoHAR* (Our)**	**0.79 ± 0.0207**	**0.81 ± 0.0176**	**0.75 ± 0.0248**	**0.79 ± 0.0117**	0.82 ± (0.0069)
Average Improvement	**≈ 0.0389**	**≈ 0.0409**	**≈ 0.0197**	**≈ 0.0416**	**≈ 0.0502**

(b) Lower (source) to Upper (target)

these results demonstrate that *CoHAR* not only improves feature alignment across users but also preserves strong class discriminability under real-world variability.

4.3.2 Robustness Analysis of *CoHAR* (RQ2)

We aim to assess the robustness and scalability of our proposed framework, *CoHAR*, by comparing it with various baseline algorithms to handle cross-gender variations.



(a) Augmentation on target domain

(b) Augmentation on source domain

(c) Augmentation on source and target domains

Fig. 8: The ablation study highlighting the performance difference between **(A), (B), (C), (D)** corresponds to *Baseline [4]*, *Temporal Ensembling*, *Temporal Ensembling + kCMMD loss* and *Temporal Ensembling + kCMMD loss + Consistency Regularizer*, respectively across all the benchmark datasets.

The experiment involves training the models using data from male subjects (source users) and evaluating their performance on data from female subjects (target users). The results are highlighted as shown in Tables 2 and 3, indicate that *CoHAR*'s significantly outperforms other benchmarks, achieving an **average macro F1 score** improvement of **[5.3 ± (2.14)]%** for both the experiment setup. *CoHAR*'s superiority can be attributed to its effective use of temporal ensembling, where the α parameter plays a crucial role in discerning unique instance-based learning features across genders. Additionally, incorporating the kCMMD loss and consistency regularization is instrumental in capturing non-linear features within a high-dimensional space and providing consistent predictions. Further analysis on optimizing the α parameter and its impact on pseudo label quality is detailed in section 4.3.4.

Another noteworthy finding is the strong performance of ContrasGAN and CoTMix in cross-gender UDA across two scenarios, ranking second among the baseline algorithms. Due to the bi-directional GAN and class-specific alignment through contrastive learning, mean maximum discrepancy, and discrepancy loss, enabling it to effectively capture and learn the distinct domain-invariant features across different users and similarly by employing a contrastive learning approach to learn the cross-domain via temporal mixup to build a robust feature embedding capturing distinct patterns for the ContrasGAN and CoTMix algorithms, respectively. Additionally, we present a comparative analysis of cross-dexterity user adaptation recognition performance across the best SOTA algorithm and *CoHAR*employing the MM-DOS dataset to highlight the performance. We plot confusion matrices, depicted in Figure 7. Notably, we observe an approximate improvement of **11%** between the proposed *CoHAR*framework and the SWL-Adapt approach.

Similarly, in tables 4 summarizes the cross-body adaptation performance across state-of-the-art (SOTA) benchmark datasets under two scenarios: (a) *upper body (source) to lower body (target)*, and (b) *lower body (source) to upper body (target)*. The results show that ***CoHAR*** consistently outperforms existing SOTA UDA algorithms, achieving an average improvement of [1-5%]. This demonstrates the effectiveness of the

proposed framework in generalizing across heterogeneous sensor positions, addressing domain distribution shifts, and learning robust domain-invariant features. SWL-Adapt ranks as the second-best performance across several datasets via leveraging meta-optimization and pseudo-labeling strategies to improve domain discrimination. Furthermore, these results underscore that *CoHAR*is highly effective in various adaptation settings, including scenarios of cross-user/-gender/-expertise, outperforming SOTA methods in the wHAR domain.

4.3.3 Comparative Analysis of *CoHAR* (RQ3)

This analysis aims to discern the individual and combined impacts of temporal ensembling (TE), kCMMD, and consistency regularization on *CoHAR*performance. We compute the average macro F1 scores for four methods of cross-domain adaptation across three different augmentation settings highlighted in Figure 8. Furthermore, our findings demonstrate a significant improvement when incorporating all methods for cross-domain adaptation, improvement ranging from [**22-35%**] across three augmentation settings shown in Figures 8a, 8b, and 8c. We examined an interesting notion: when performing augmentation on the target domain, the performance notably decreases compared to the other augmentation settings. One of the primary reasons might be the generation of noisy pseudo labels due to the variations only occurring in the target domain, whereas when augmentation is performed on the source domain, the performance is comparable to that of augmentation performed on both domains.

4.3.4 Compatibility Analysis of *CoHAR* (RQ4)

In our study, we investigated the effectiveness of the self-ensembling technique in generating reliable pseudo-labels for the target domain. We utilized **entropy**, $H(x) = -\sum_{i=1}^{n} p(x_i) \log(p(x_i))$, as a measure to quantify the uncertainty of model predictions. As depicted in Figure 9, there is an inverse relationship between the α parameter of ensembling and the entropy values; a higher α results in lower uncertainty, thus a reduction in entropy indicates that the model becomes increasingly certain about its predictions, i.e. generating more stable pseudo-labels. Our experiments indicate that an α range of [0.55-0.75] is optimal for producing stable, high-quality pseudo-labels. This observation is further validated in Figure 9, where we demonstrate that an α value of **0.60** achieves faster convergence compared to other α values. This insight underscores the pivotal role of the α parameter in temporal ensembling for enhancing the quality of pseudo-labels in domain adaptation tasks.

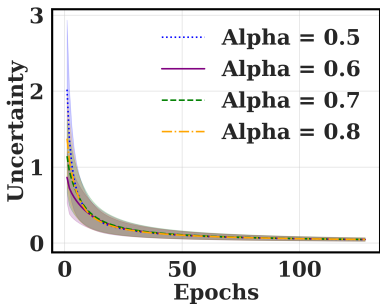


Fig. 9: Correlation between α parameter and uncertainty.

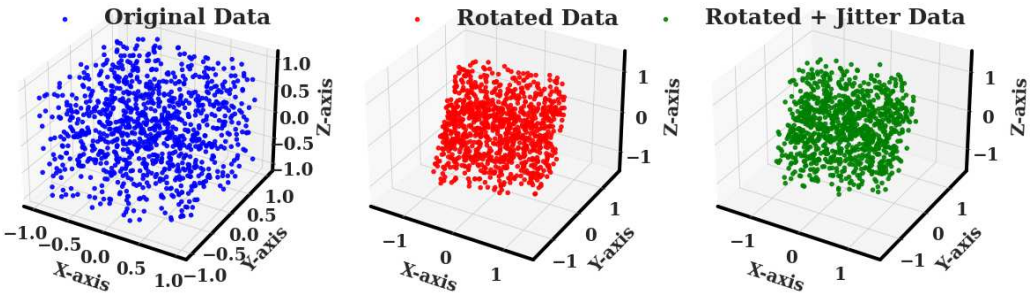
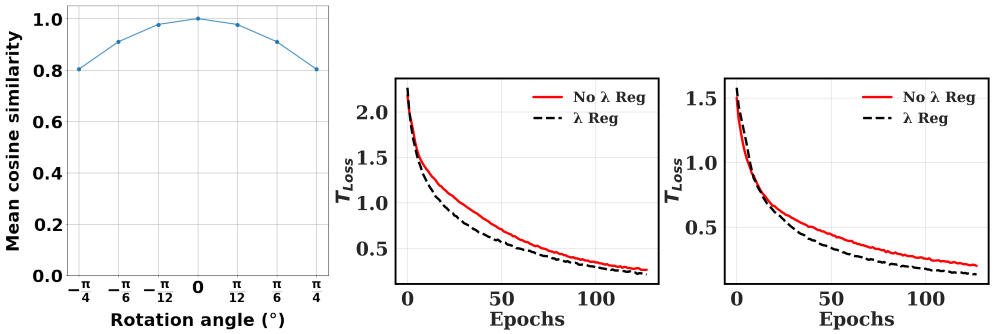


Fig. 10: Example for Geometric-based Augmentation Samples



(a) Cosine similarity under rotation. (b) Impact of λ (BAR). (c) Impact of λ (MMDOS).

Fig. 11: (a) Cosine similarity between real and augmented samples across varying rotation angles for BAR dataset. (b-c) Effect of λ regularization weight on training loss for BAR and MMDOS datasets, respectively.

4.3.5 Effectiveness Analysis of *CoHAR* (RQ5)

We conducted an effectiveness analysis of *CoHAR*to evaluate the effectiveness of data augmentation in generating stable samples. This analysis also examines the role of data augmentation in enhancing model robustness, as well as the impact of the regularization variable, λ, on the overall performance of *CoHAR*on domain adaptation as discussed in Equation 8.

- Impact of Data Augmentation on *CoHAR's* Generalizability: We demonstrate how the augmentation module generates robust representation samples while preserving the underlying data distribution and label consistency. The augmented samples adhere to the three principles outlined in Section 3. We applied rotation within the range $[-\pi/4$ ¡-¿ $+\pi/4]$, enhancing the model's robustness to spatial variations. Specifically, rotation magnitudes between $r = 5°$ and $20°$ yielded the best results

across benchmark datasets, effectively balancing augmentation diversity with preserving underlying data distribution. In addition, Gaussian noise-based jitter with a standard deviation (σ) in the range 0.01 to 0.10 was introduced to provide controlled variability. Jitter values between 0.03 and 0.06 consistently delivered optimal results across datasets, ensuring realistic, diverse, and label-preserving augmented samples. These geometric-based augmentation strategies enhance *CoHAR's* ability to model invariances inherent in temporal wearable data, effectively capturing the real-world variations prevalent in wHAR tasks. In Figures 10, we highlight examples of data augmentation: (a) corresponds to ground-truth samples, (b) represents a rotated sample by $\pi/4$, and (c) shows a sample rotated and combined with jitter (SD = 0.05). These examples highlight the module's ability to preserve label consistency while injecting controlled variability that mirrors real-world conditions. To quantify the effect, we compute the *mean cosine similarity* between each original and rotated sample (Fig. 11a) and also across the rotation interval $r_\theta \in [-\pi/4, \pi/4]$, similarity scores never drop below 0.80 and remain ≥ 0.95 for rotations up to $\pm\pi/6$. These results indicate that the augmentation introduces controlled diversity while *likely* preserving label fidelity and avoiding substantial out-of-distribution (OOD) shifts.

- Impact of λ on Conditional Domain Alignment: The impact of the regularization parameter λ on the total loss T_{Loss} is demonstrated in Figures 11b and 11c. The red curve corresponds to the scenario without λ-based regularization, where the total loss decreases steadily but exhibits a relatively uncontrolled convergence in both scenarios. This behavior indicates a risk of overfitting, as the optimization process lacks a mechanism to constrain the learned representations effectively. Furthermore, in contrast, the black dashed curve incorporates λ regularization, which stabilizes training dynamics by conditioning the kernel matrices K_{ss} and K_{tt} with the regularization term λI. This mitigates ill-conditioning and overfitting while balancing intra-domain consistency and inter-domain alignment. By minimizing the conditional probability gap between the source ($P^s(Z|Y)$) and target ($P^t(Z|Y)$) distributions within the RKHS embedding, λ ensures a robust, domain-invariant feature space showcased improved optimization stability and improved generalization across domains.

5 Limitation and Future Work

Despite the strong performance of *CoHAR* across diverse label spaces and variations in state-of-the-art wHAR datasets, we acknowledge some limitations that can be addressed and improved with additional detail and refinement. The adaptive augmentation strategy relies on predefined geometric and frequency perturbations, which may not adequately capture real-world out-of-distribution (OOD) scenarios [65, 66]. Moreover, the augmentation module lacks input-aware adaptability, which limits its robustness to unseen/varied real-world settings. Additionally, the kernel-based kCMMD alignment and iterative pseudo-labeling steps introduce notable computational overhead, particularly when dealing with long sequences or large-scale

deployments. In the future, we plan to expand the capabilities of the *CoHAR* framework by focusing on multimodal sensor fusion, for instance, IMU and audio, and improved adaptation for sensors positioned across the body. Furthermore, we plan to expand the capabilities of the *CoHAR* framework by exploring higher-order statistics-based methods, such as cross-covariance estimation for conditional embedding alignment. These enhancements will advance domain generalization in the wHAR domain, paving the way for more versatile and effective action recognition systems.

6 Conclusion

This work introduced a novel single-stage joint optimization strategy for cross-user wearable Human Action Recognition (wHAR). Our approach integrates *temporal-ensembling* with *kernel-based class-wise conditional mean maximum discrepancy* (*kCMMD*), effectively generating high-quality pseudo labels and minimizing conditional domain discrepancies, respectively. Additionally, consistency regularization enforced via KL divergence promotes stable label predictions across augmented samples, thereby improving robustness to perturbations induced by inter-user variability, sensor noise, and domain shifts. The *CoHAR* framework exhibited notable superiority over benchmark UDA algorithms in extensive experiments, demonstrating its generalizability and robustness, outperforming SOTA approaches with a **4-12%** improvement.

7 Acknowledgment

This work has been partially supported by NSF CNS EAGER Grant #2233879, U.S. Army Grants #W911NF2120076 and #W911NF2410367, NSF IIS Grant #2509680, ONR Grant #N00014-23-1-2119, NSF CAREER Award #1750936, and NSF REU Site Grant #2050999.

References

[1] Ramamurthy, S.R., Ghosh, I., Gangopadhyay, A., Galik, E., Roy, N.: Star: A scalable self-taught learning framework for older adults' activity recognition. In: 2021 IEEE International Conference on Smart Computing (SMARTCOMP), pp. 121–128 (2021). IEEE

[2] Serpush, F., Menhaj, M.B., Masoumi, B., Karasfi, B.: Wearable sensor-based human activity recognition in the smart healthcare system. Computational intelligence and neuroscience **2022**(1), 1391906 (2022)

[3] Ghosh, I., Chakma, A., Ramamurthy, S.R., Roy, N., Waytowich, N.: Permtl: A multi-task learning framework for skilled human performance assessment. In: 2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 37–44 (2022). IEEE

[4] Ghosh, I., Ramamurthy, S.R., Chakma, A., Roy, N.: Decoach: Deep learning-based coaching for badminton player assessment. Pervasive and Mobile Computing, 101608 (2022)

[5] Xie, J., Wen, D., Liang, L., Jia, Y., Gao, L., Lei, J., *et al.*: Evaluating the validity of current mainstream wearable devices in fitness tracking under various physical activities: comparative study. JMIR mHealth and uHealth **6**(4), 9754 (2018)

[6] Phukan, N., Mohine, S., Mondal, A., Manikandan, M.S., Pachori, R.B.: Convolutional neural network-based human activity recognition for edge fitness and context-aware health monitoring devices. IEEE Sensors Journal **22**(22), 21816–21826 (2022)

[7] Zhang, X., Yao, L., Wang, X., Zhang, W., Zhang, S., Liu, Y.: Know your mind: Adaptive cognitive activity recognition with reinforced cnn. In: 2019 IEEE International Conference on Data Mining (ICDM), pp. 896–905 (2019). IEEE

[8] Villani, V., Righi, M., Sabattini, L., Secchi, C.: Wearable devices for the assessment of cognitive effort for human–robot interaction. IEEE Sensors Journal **20**(21), 13047–13056 (2020)

[9] Sano, A., Taylor, S., McHill, A.W., Phillips, A.J., Barger, L.K., Klerman, E., Picard, R.: Identifying objective physiological markers and modifiable behaviors for self-reported stress and mental health status using wearable sensors and mobile phones: observational study. Journal of medical Internet research **20**(6), 210 (2018)

[10] Zhu, C., Sheng, W.: Wearable sensor-based hand gesture and daily activity recognition for robot-assisted living. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans **41**(3), 569–573 (2011)

[11] Saleem, G., Bajwa, U.I., Raza, R.H.: Toward human activity recognition: a survey. Neural Computing and Applications **35**(5), 4145–4182 (2023)

[12] Li, C., Tong, C.L., Niu, D., Jiang, B., Zuo, X., Cheng, L., Xiong, J., Yang, J.: Similarity embedding networks for robust human activity recognition. ACM Transactions on Knowledge Discovery from Data (TKDD) **15**(6), 1–17 (2021)

[13] Sun, X., Kashima, H., Tomioka, R., Ueda, N., Li, P.: A new multi-task learning method for personalized activity recognition. In: 2011 IEEE 11th International Conference on Data Mining, pp. 1218–1223 (2011). IEEE

[14] Liu, X., Yoo, C., Xing, F., Oh, H., El Fakhri, G., Kang, J.-W., Woo, J., et al.: Deep unsupervised domain adaptation: A review of recent advances and perspectives. APSIPA Transactions on Signal and Information Processing **11**(1) (2022)

[15] Ige, A.O., Noor, M.H.M.: A survey on unsupervised learning for wearable sensor-based activity recognition. Applied Soft Computing **127**, 109363 (2022)

[16] Chang, Y., Mathur, A., Isopoussu, A., Song, J., Kawsar, F.: A systematic study of unsupervised domain adaptation for robust human-activity recognition. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies **4**(1), 1–30 (2020)

[17] Hosseini, A., Zamanzadeh, D., Valencia, L., Habre, R., Bui, A.A., Sarrafzadeh, M.: Domain adaptation in children activity recognition. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 1725–1728 (2019). IEEE

[18] Ghosh, I., Chugh, G., Md Faridee, A.Z., Roy, N.: Unsupervised domain adaptation for action recognition via self-ensembling and conditional embedding alignment. In: 2024 IEEE International Conference on Data Mining (ICDM), pp. 699–704 (2024). https://doi.org/10.1109/ICDM59182.2024.00079

[19] Akbari, A., Jafari, R.: Transferring activity recognition models for new wearable sensors with deep generative domain adaptation. In: Proceedings of the 18th International Conference on Information Processing in Sensor Networks, pp. 85–96 (2019)

[20] Zhang, Y., Bin, M., Zhang, Y., Wang, Z., Han, Z., Liang, C.: Link-based contrastive learning for one-shot unsupervised domain adaptation. In: Proceedings of the Computer Vision and Pattern Recognition Conference, pp. 4916–4926 (2025)

[21] Xu, X., Hu, Y., Tang, H., Qin, Y., Mi, L., Wang, H., Li, X.: Concept-based unsupervised domain adaptation. arXiv preprint arXiv:2505.05195 (2025)

[22] Barbosa, P., Garcia, K.D., Mendes-Moreira, J., Carvalho, A.C.: Unsupervised domain adaptation for human activity recognition. In: Intelligent Data Engineering and Automated Learning–IDEAL 2018: 19th International Conference, Madrid, Spain, November 21–23, 2018, Proceedings, Part I 19, pp. 623–630 (2018). Springer

[23] Zhou, Z., Zhang, Y., Yu, X., Yang, P., Li, X.-Y., Zhao, J., Zhou, H.: Xhar: Deep domain adaptation for human activity recognition with smart devices. In: 2020 17th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON), pp. 1–9 (2020). IEEE

[24] Combes, R., Zhao, H., Wang, Y.-X., Gordon, G.J.: Domain adaptation with conditional distribution matching and generalized label shift. Advances in Neural Information Processing Systems **33**, 19276–19289 (2020)

[25] Stikic, M., Van Laerhoven, K., Schiele, B.: Exploring semi-supervised and active learning for activity recognition. In: 2008 12th IEEE International Symposium

on Wearable Computers, pp. 81–88 (2008). IEEE

[26] Hu, R., Chen, L., Miao, S., Tang, X.: Swl-adapt: An unsupervised domain adaptation model with sample weight learning for cross-user wearable human activity recognition. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 37, pp. 6012–6020 (2023)

[27] Sanabria, A.R., Zambonelli, F., Dobson, S., Ye, J.: Contrasgan: Unsupervised domain adaptation in human activity recognition via adversarial and contrastive learning. Pervasive and Mobile Computing **78**, 101477 (2021)

[28] Eldele, E., Ragab, M., Chen, Z., Wu, M., Kwoh, C.-K., Li, X.: Cotmix: Contrastive domain adaptation for time-series via temporal mixup. arXiv preprint arXiv:2212.01555 (2022)

[29] Liu, Q., Xue, H.: Adversarial spectral kernel matching for unsupervised time series domain adaptation. In: IJCAI, pp. 2744–2750 (2021)

[30] Chen, C., Fu, Z., Chen, Z., Jin, S., Cheng, Z., Jin, X., Hua, X.-S.: Homm: Higher-order moment matching for unsupervised domain adaptation. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 3422–3429 (2020)

[31] Mirza, M.J., Micorek, J., Possegger, H., Bischof, H.: The norm must go on: Dynamic unsupervised domain adaptation by normalization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 14765–14775 (2022)

[32] Altun, K., Barshan, B.: Human activity recognition using inertial/magnetic sensor units. In: International Workshop on Human Behavior Understanding, pp. 38–51 (2010). Springer

[33] Reiss, A., Stricker, D.: Creating and benchmarking a new dataset for physical activity monitoring. In: Proceedings of the 5th International Conference on PErvasive Technologies Related to Assistive Environments, pp. 1–8 (2012)

[34] Sharshar, A., Fayez, A., Eitta, A.A., Gomaa, W.: Mm-dos: A novel dataset of workout activities. In: 2022 International Joint Conference on Neural Networks (IJCNN), pp. 1–8 (2022). IEEE

[35] Ghosh, I., Goldstein, A., Chakma, A., Freeman, J., Gregory, T., Suri, N., Ramamurthy, S.R., Roy, N.: Heterosys: Heterogeneous and collaborative sensing in the wild. In: 2023 IEEE International Conference on Smart Computing (SMARTCOMP), pp. 285–290 (2023). IEEE

[36] Ebner, C.J., Findling, R.D.: Tennis stroke classification: Comparing wrist and racket as imu sensor position. In: Proceedings of the 17th International Conference on Advances in Mobile Computing & Multimedia, pp. 74–83 (2019)

[37] Gowda, M., Dhekne, A., Shen, S., Choudhury, R.R., Yang, L., Golwalkar, S., Essanian, A.: Bringing iot to sports analytics. In: 14th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 17), pp. 499–513 (2017)

[38] Li, S., Zhang, W.: Evaluation method of basketball teaching and training effect based on wearable device. Frontiers in Physics **10**, 372 (2022)

[39] Radhakrishnan, M., Misra, A., Balan, R.K.: W8-scope: Fine-grained, practical monitoring of weight stack-based exercises. Pervasive and Mobile Computing **75**, 101418 (2021)

[40] Khalifa, S., Lan, G., Hassan, M., Seneviratne, A., Das, S.K.: Harke: Human activity recognition from kinetic energy harvesting data in wearable devices. IEEE Transactions on Mobile Computing **17**(6), 1353–1368 (2017)

[41] Peralta, D., Van Herbruggen, B., Fontaine, J., Debyser, W., Wieme, J., De Poorter, E.: Badminton stroke classification based on accelerometer data: from individual to generalized models. In: 2022 IEEE International Conference on Big Data (Big Data), pp. 5542–5548 (2022). IEEE

[42] Haresamudram, H., Essa, I., Plötz, T.: Contrastive predictive coding for human activity recognition. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies **5**(2), 1–26 (2021)

[43] Tian, J., Zhou, P., Sun, F., Wang, T., Zhang, H.: Wearable imu-based gym exercise recognition using data fusion methods. In: The Fifth International Conference on Biological Information and Biomedical Engineering, pp. 1–7 (2021)

[44] Um, T.T., Babakeshizadeh, V., Kulić, D.: Exercise motion classification from large-scale wearable sensor data using convolutional neural networks. In: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 2385–2390 (2017). IEEE

[45] Qin, X., Wang, J., Ma, S., Lu, W., Zhu, Y., Xie, X., Chen, Y.: Generalizable low-resource activity recognition with diverse and discriminative representation learning. In: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery, New York, NY, USA (2023). https://doi.org/10.1145/3580305.3599360 . https://doi.org/10.1145/3580305.3599360

[46] Hossain, H.S., Roy, N.: Active deep learning for activity recognition with context aware annotator selection. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 1862–1870 (2019)

[47] Ghosh, I., Chugh, G., Faridee, A.Z.M., Roy, N.: High-order moments conditional domain adaptation networks for wearable human activity recognition. In: Proceedings of the 34th ACM International Conference on Information and Knowledge Management, pp. 718–728 (2025)

[48] Yang, S., Dong, Y., Ward, R., Dhillon, I.S., Sanghavi, S., Lei, Q.: Sample efficiency of data augmentation consistency regularization. In: International Conference on Artificial Intelligence and Statistics, pp. 3825–3853 (2023). PMLR

[49] Zha, Q., Liu, X., Peng, S.-J., Cheung, Y.-m., Xu, X., Wang, N.: Recon: Enhancing true correspondence discrimination through relation consistency for robust noisy correspondence learning. In: Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR), pp. 29680–29689 (2025)

[50] Wang, L., Xu, L., Yang, X., Huang, Z., Cheng, J.: Debiased distillation for consistency regularization. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 39, pp. 7799–7807 (2025)

[51] Englesson, E., Azizpour, H.: Consistency regularization can improve robustness to label noise. arXiv preprint arXiv:2110.01242 (2021)

[52] Xie, Q., Dai, Z., Hovy, E., Luong, T., Le, Q.: Unsupervised data augmentation for consistency training. Advances in neural information processing systems **33**, 6256–6268 (2020)

[53] Lowell, D., Howard, B.E., Lipton, Z.C., Wallace, B.C.: Unsupervised data augmentation with naive augmentation and without unlabeled data. arXiv preprint arXiv:2010.11966 (2020)

[54] Fan, Y., Kukleva, A., Dai, D., Schiele, B.: Revisiting consistency regularization for semi-supervised learning. International Journal of Computer Vision **131**(3), 626–643 (2023)

[55] Um, T.T., Pfister, F.M., Pichler, D., Endo, S., Lang, M., Hirche, S., Fietzek, U., Kulić, D.: Data augmentation of wearable sensor data for parkinson's disease monitoring using convolutional neural networks. In: Proceedings of the 19th ACM International Conference on Multimodal Interaction, pp. 216–220 (2017)

[56] Laine, S., Aila, T.: Temporal ensembling for semi-supervised learning. arXiv preprint arXiv:1610.02242 (2016)

[57] Oliveira, W.D.G., Berton, L.: A systematic review for class-imbalance in semi-supervised learning. Artificial Intelligence Review **56**(Suppl 2), 2349–2382 (2023)

[58] Ren, Y., Zhu, J., Li, J., Luo, Y.: Conditional generative moment-matching networks. Advances in Neural Information Processing Systems **29** (2016)

[59] Ren, C.-X., Ge, P., Dai, D.-Q., Yan, H.: Learning kernel for conditional moment-matching discrepancy-based image classification. IEEE transactions on cybernetics **51**(4), 2006–2018 (2019)

[60] Batista, G.E., Wang, X., Keogh, E.J.: A complexity-invariant distance measure for time series. In: Proceedings of the 2011 SIAM International Conference on Data Mining, pp. 699–710 (2011). SIAM

[61] Faridee, A.Z.M., Khan, M.A.A.H., Pathak, N., Roy, N.: Augtoact: Scaling complex human activity recognition with few labels. In: Proceedings of the 16th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, pp. 162–171 (2019)

[62] Gouk, H., Frank, E., Pfahringer, B., Cree, M.J.: Regularisation of neural networks by enforcing lipschitz continuity. Machine Learning **110**, 393–416 (2021)

[63] Pereira, D.G., Afonso, A., Medeiros, F.M.: Overview of friedman's test and post-hoc analysis. Communications in Statistics-Simulation and Computation **44**(10), 2636–2653 (2015)

[64] Ben-David, S., Blitzer, J., Crammer, K., Kulesza, A., Pereira, F., Vaughan, J.W.: A theory of learning from different domains. Machine learning **79**(1), 151–175 (2010)

[65] Sun, M., Li, Y., Wang, Y., Wang, X.: Towards domain-aware stable meta learning for out-of-distribution generalization. ACM Transactions on Knowledge Discovery from Data **18**(8), 1–24 (2024)

[66] Yang, Y., Wei, H., Sun, Z.-Q., Li, G.-Y., Zhou, Y., Xiong, H., Yang, J.: S2osc: A holistic semi-supervised approach for open set classification. ACM Transactions on Knowledge Discovery from Data (TKDD) **16**(2), 1–27 (2021)