Unsupervised Domain Adaptation for Action Recognition via Self-Ensembling and Conditional Embedding Alignment

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

[1]Department of Information Systems, University of Maryland, Baltimore County, United States
[2]Department of Computer Science and Engineering, Indian Institute of Technology Jodhpur, India
[3]Amazon, United States
[1](indrajeetghosh, nroy)@umbc.edu, [2]chugh.2@iitj.ac.in, [3]abufari@amazon.com

Abstract—Recent advancements in deep learning-based wearable human action recognition (wHAR) have improved the capture and classification of complex motions, but 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 *μDAR*, 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 (kCMMD) 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 *μDAR* results in a range of $\approx$ *4-12%* average macro-F1 score improvement over six state-of-the-art UDA methods in four benchmark wHAR datasets.

Index Terms—Human Activity Recognition, Consistency Regularization, Temporal Ensembling, Cross-user Adaptation

I. INTRODUCTION

Over the past decade, wearable human activity recognition (wHAR) has enabled applications in healthcare [16], sports analytics [7], and fitness tracking [26], leveraging inertial, gyroscopic, and magnetometric sensors in smartphones and smartwatches to capture motion data, with neural processing units facilitating real-time inference. However, traditional models rely on supervised learning and extensive labeled data, limiting scalable, robust activity recognition. While Activities of Daily Living (ADLs) datasets cover simple actions like *running* and *walking*, micro-complex activities show user-specific patterns, leading to cross-user variations [23] due to differences in body characteristics and skill levels [8]. These variations are particularly evident in skill-based actions, where proficiency evolves; for example, experienced badminton players exhibit more consistent patterns than novices [7], reflecting how skill progression influences activity patterns.

In this context, unsupervised domain adaptation (UDA) [14] offers a viable solution, particularly when annotations are limited and domain discrepancies arise from out-of-distribution samples. UDA leverages labeled data from the source domain (e.g., female, expert player) and unlabeled data from the target domain (e.g., male, novice player), aligning feature distributions to learn a domain-invariant feature space. Most UDA research [2] focuses on marginal feature alignment, with few models addressing conditional distribution alignment, often requiring labeled target data [24]. Pseudo-labeling can be an alternative, but its self-labeling nature often amplifies initial classification errors [22], hindering effective conditional alignment across domains.

To address these challenges, we introduce a novel joint optimization framework using complementary loss components. We begin by training a deep learning classifier on labeled source data. Next, we adapt this model to unlabeled target data by learning a domain-invariant embedding via minimizing the class-wise kernel-based conditional maximum mean discrepancy (kCMMD) distance, using pseudo-labels from the target classifier. To prevent training collapse due to incorrect pseudo-labels, we employ a temporal ensemble approach to smooth them via aggregation. To enhance generalizability, we apply unsupervised data augmentation to both source and target data, coupled with consistency training, by minimizing the KL divergence between predictions of original and augmented samples. This ensures that augmented samples receive consistent labels, increasing robustness to data noise and covering uncertain regions in the decision space. It also maintains consistent pseudo-labels for target samples across augmentations, which, along with the temporal ensemble, provides high-quality pseudo-labels for conditional kernel regularization. This synergy minimizes the domain gap

and improves target domain classification without labeled samples. Consequently, *μDAR* outperforms state-of-the-art (SOTA) UDA baselines on four activity recognition datasets, covering both simple daily activities and complex sports/gym actions.

The key contributions are summarized as follows:

- We propose a novel UDA framework (*μDAR*) to address two challenges in wHAR: (i) *cross-user variations* and (ii) *limited labeled data*. Our method learns source-specific representations from labeled data and applies them to unlabeled target data. We open-source *μDAR*[1] for reproducibility.

- We develop a novel joint optimization architecture incorporating: (i) *kernel-based class-wise conditional maximum mean discrepancy* (*k*CMMD) to align conditional distributions between source and target features; (ii) a *temporal ensemble-based pseudo-labeling* technique for robust target labels; and (iii) a *consistency regularizer between augmentations and real samples* to enhance generalizability and robustness, improving pseudo-labels of target domain. We detail the synergy of these components via ablation studies.

- We compare *μDAR's* macro-F1 performance against six SOTA UDA algorithms [10, 20, 5, 13, 3, 15] on four public benchmark datasets [8, 1, 17, 21] covering varying label complexities, dexterity levels (novice vs. expert). *μDAR* consistently outperforms baselines by $\approx$ **4–12%** over five independent trials per dataset.

II. RELATED WORK

We discuss the relevant literature to *μDAR* on UDA in the wHAR and consistency regularization area with an emphasis on handling cross-user variations and limited labeled data and to differentiate *μDAR* from the SOTA approaches.

1) Wearable-based Unsupervised Domain Adaptation (UDA): have been significant advances in wearable-based recognition of daily human activities [11], 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. UDA has become a valuable method for overcoming the challenge of limited labelled 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 [10], 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. However, recent SOTA studies have addressed minimal label information by minimizing conditional feature distributions. In contrast, this work focuses on tackling limited label data by minimizing conditional discrepancies.

2) Consistency Regularization: has recently become prominent, especially in developing robust models invariant to data variations, i.e., augmented samples [28]. The core principle
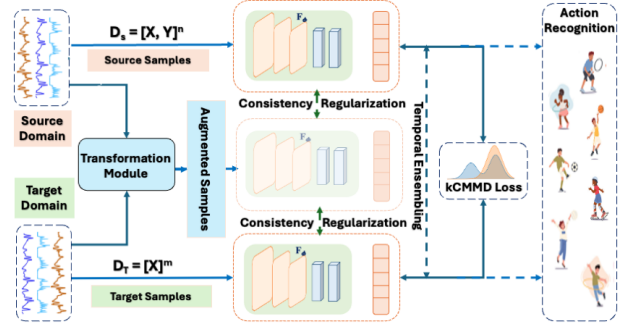
Fig. 1: High level overview of the *μDAR* architecture

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 [6]. Studies [27] indicate that combining consistency regularization with unsupervised data augmentation in semi-supervised learning impedes performance degradation and bolsters model generalizability and robustness. Motivating by this, we integrate a consistent regularization approach to tackle user variations. However, to design real-world settings, we incorporate unsupervised data augmentation for wHAR [25], and to our best knowledge, our work is among the first few that have integrated unsupervised data augmentation with consistency regularization in wHAR.

III. METHODOLOGY

In this section, we present the *μDAR* framework for the wHAR domain, designed to recognize ADL actions, including gym and sports activities. We briefly outline the problem formulation and highlight the key functional and learning components of *μDAR*.

A. Problem Formulation

Cross-user variations and limited expert-labeled datasets hinder the effectiveness of wHAR methods. We leverage labeled data from experienced users (source domain), $D_s = \{(x_i, y_i)\}_{i=1}^n$, and unlabeled data from new users (target domain), $D_t = \{x_j\}_{j=1}^m$. Here, n and m represent the number of instances in the source and target domains, respectively, with both sharing the same label space, $y_s^i, y_t^j \in \{1, 2, \ldots, c\}$, where c denotes the number of activity classes. UDA aims to identify activities for new users by utilizing labeled source data alongside unlabeled target data. This assumes that both domains use identical sensors and placements, performing the same activities, despite conditional distribution shifts, i.e., $P(y \mid x_s) \neq P(y \mid x_t)$.

Figure 1 illustrates the proposed UDA framework, comprising four key components to tackle these challenges: (i) learning source-specific features to capture trained user characteristics; (ii) generating high-quality pseudo labels via temporal ensembling; (iii) aligning source and target features using the proposed *k*CMMD loss; and (iv) applying unsupervised data

augmentation to improve robustness against user-variations. This integrated approach minimizes domain discrepancies, enhancing generalization to new users in the wHAR domain.

B. 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 **CE supervised loss** $\mathcal{L}_{SL} = -\frac{1}{n} \sum_{i=1}^{n} P(y_{s_i}|x_{s_i}) \log P(\hat{y}_{s_i}|x_{s_i})\}$ and ensure accurately predict the class distribution $P(\hat{Y}|X)$.

C. 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 activities' inherent variability and complexity, 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* [12]. This method is designed to maximize the concurrent extraction of mutual information from both the source and target domains, which aims to enhance the pseudo-label generation process by aggregating model predictions in an iterative refinement update method highlighted in Eq. 1, where $\mathcal{L}$, θ, η, t and ∇_θ represents loss, model parameters, learning rate, iteration step and gradient update, respectively. This aggregation smoothens the overall label prediction and enhances the quality of the pseudo labels for the target domain.

Temporal ensembling is a robust mechanism for identifying and learning domain-invariant features, thereby reducing discrepancies between the source and target data over time. The model's adaptation to the target domain improves significantly by averaging predictions over time. Temporal ensembling mitigates class imbalance [4] by generating pseudo-labels and enabling stable learning, which helps in learning a balanced data representation and alleviates the skew toward majority classes seen in the labeled data. We determine the optimal setting for α using a grid-search method and then fine-tuned it for the optimal performance across the datasets.

$$\theta_{t+1} = \theta_t - \eta \cdot \nabla_\theta \mathcal{L} \qquad (1)$$

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

We focus on aligning conditional distributions using our proposed kernel-based class-wise conditional mean maximum discrepancy (kCMMD) approach in the Reproducing Kernel Hilbert Space (RKHS) [19]. In RKHS, conditional distributions are mapped to a high-dimensional feature space,

and kCMMD computes the mean discrepancy between these embeddings to minimize domain shifts. This reduces the conditional probability gap between the source ($P^s(Z|Y)$) and target ($P^t(Z|Y)$) domains, establishing a robust conditional domain-invariant feature space. The generalized kCMMD loss is defined in Eq. 2, 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.

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

Considering our problem scenario, where the source domain is represented by a feature set $x_s \in \mathbb{R}^{n \times d}$ with corresponding labels y_s, and a target domain represented by $x_t \in \mathbb{R}^{m \times d}$ with pseudo-labels $y_{\bar{t}}$, the kCMMD loss focuses on minimizing distribution differences for each class $c \in C$. 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 the source (X_s) and target (X_t) domains, we compute the following kernel matrices (K_M), quantifying intra- and inter-domain discrepancies using element-wise multiplication [$\otimes$]:

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

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

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

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, control overfitting [19, 18] and class-imbalance during intra-domain computation. 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 so 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 overall regularized kCMMD discrepancy loss function for the kernel matrices K_M^{ss}, K_M^{tt}, and K_M^{st} is defined in Equation (3):

$$\mathcal{L}_{kCMMD} = (\bar{K}_M^{ss} + \bar{K}_M^{tt} - 2 \cdot \bar{K}_M^{st}) \qquad (3)$$

Equation 3 measures the overall conditional discrepancy between the source and target domain distributions. Regularized $\bar{K}_M^{ss}$ and $\bar{K}_M^{tt}$ represent the mean intra-domain similarities for the source and target domains, respectively, while $\bar{K}_M^{st}$ 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.

E. Unsupervised Data Augmentation Module

We address user variability such as differences in activity execution, proficiency levels and individual traits by integrating an unsupervised data augmentation approach [27] to build robustness against cross-user variations. Our augmentation module ensures that (i) augmented samples are valid and realistic, (ii) they exhibit diversity, and (iii) they introduce specific inductive biases. Recognizing that sudden temporal shifts in wearable data can impair model performance, we develop a model invariant with such transformations. To achieve this, we implement two geometric-based augmentations: *jitter* introduces high-frequency fluctuations in the data, distinct from actual motion, thereby adding stochastic noise and ***rotation*** which involves angular displacement or orientation change, quantifying the degree and axis of rotational movement which are well-established for capturing real-world variations in wHAR tasks [25]. Let G denote the geometric augmentation applied to the source and target data, x_s and x_t, defined as $G = [r_\theta, \sigma]$. The augmented samples are then $x'_s = G \cdot x_s$ and $x'_t = G \cdot x_t$. Our goal is to enforce consistency between the model's predictions on original and augmented data, ensuring robustness to user-specific variations. We employ a consistency loss (Eq. (4)), where $f_\theta(x_s)$, $f_\theta(x'_s)$, $f_\theta(x_t)$, and $f_\theta(x'_t)$ represent predictions for real and augmented samples. This approach aligns with the Lipschitz continuity theorem [9], ensuring the model's outputs change proportionally with input variations, enhancing generalization to diverse user variations.

$$\mathcal{L}_C = \mathbb{E}x_s \sim D_s, x'_s \sim t(x_s) \left[D_{KL}(f_\theta(x_s)||f_\theta(x'_s)) \right] + \\ \mathbb{E}x_t \sim D_t, x'_t \sim t(x_t) \left[D_{KL}(f_\theta(x_t)||f_\theta(x'_t)) \right] \quad (4)$$

F. μDAR Training Procedure

The overall aim is to obtain a set of optimized model parameters trained on labeled source data that work well on 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 μDAR framework enumerated in Equation (5). In the overall formulation, θ encompasses all the parameters within the deep network. The hyperparameters β_0 and β_1 are set to 1, facilitating the fine-tuning of the overall objective function.

$$\min_{F_\theta} \mathcal{L}_{overall} = \underbrace{\mathcal{L}_{SL} + \beta_0 \mathcal{L}_{kCMMD}}_{\text{Domain Adaptation}} + \underbrace{\beta_1 \mathcal{L}_C}_{\text{Consistency Regularizer}}$$
$$(5)$$

TABLE I: List of *μDAR's* Hyperparameters

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 value	$5°$ - $45°$

IV. EXPERIMENTS

We evaluate μDAR via extensive experiments on the problem of limited supervision settings for ADLs action recognition by investigating the following research questions (RQs): (i) **RQ1 (Accuracy):** How well is μDAR performing on publically available datasets? (ii) **RQ2 (Robustness):** How can we adapt μDAR for accurate recognition across different users' proficiency levels in wHAR? (iii) **RQ3 (Compatibility):** Is μDAR optimized for high quality pseudo-labels generation? (iv) **RQ4 (Comparative Analysis):** Is μDAR minimizing conditional data distribution discrepancies?

A. Setup

1) Datasets: We selected **four** publicly available datasets (1) **Badminton Activity Recognition (BAR) Dataset [8]**, capturing 12 badminton strokes from 11 subjects in controlled and uncontrolled environments, with 30 iterations per stroke. (2) **Daily and Sports Activity (DSADS) Dataset [1]**, featuring 9-DOF sensor data from eight individuals performing 19 activities, recorded at 25 Hz. (3) **Physical Activity Monitoring (PAMAP2) Dataset [17]**, consisting of 18 activities from nine subjects, captured at 100 Hz from three sensors. (4) **Workout Activities (MM-DOS) Dataset [21]**, with IMU data from 50 participants performing four gym workouts, recorded at 50 Hz across nine body locations.

2) Baselines: We evaluate and contrast the performance of μDAR with **six** SOTA UDA algorithms including ***SWL-Adapt*** [10], ***ContrasGAN*** [20], ***AdvSVM*** [13], ***HoMM*** [3], ***DUA*** [15] and ***CoTMix*** [5].

B. Implementation Details

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 motion artifacts, we applied a median filter and normalized 48 features using a min-max scaler for the *BAR* dataset, which includes data from three sensor axes. Due to signal clipping beyond $\pm 2g$ in the low-noise accelerometer, we used wide/high-range accelerometers to capture rapid movements more effectively. For feature extraction, we applied a sliding window technique with varying overlap and sampling rates across the *BAR*, *PAMAP2*, *DSADS*, and *MMDOS* datasets to

TABLE II: Comparison of μDAR's Macro F1 Score ($\pm$ SD) and Average Improvement Margin vs. SWL-Adapt [10], ContrasGAN [20], CoTMix [5], HoMM [3], DUA [15] and AdvSKM [13] across BAR, DSADS, PAMAP2, and MMDOS datasets. The top and 2nd best results are in bold and underlined, respectively.

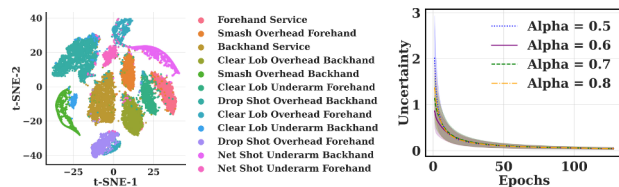| Datasets	SWL-Adapt [10]	ContrasGAN [20]	CoTMix [5]	AdvSKM [13]	HoMM [3]	DUA [15]	μDAR (Our)	Average Improvement Margin
DSADS [1]	0.70 $\pm$ (0.0188)	0.64 $\pm$ (0.0213)	0.61 $\pm$ (0.0252)	0.71 $\pm$ (0.0132)	0.68 $\pm$ (0.0265)	0.73 $\pm$ (0.0105)	**0.77 $\pm$ (0.0239)**	0.048 $\pm$ (0.015)
PAMAP2 [17]	0.73 $\pm$ (0.0174)	0.72 $\pm$ (0.0160)	0.66 $\pm$ (0.0248)	0.70 $\pm$ (0.0213)	0.67 $\pm$ (0.0229)	0.72 $\pm$ (0.0095)	**0.78 $\pm$ (0.0211)**	0.054 $\pm$ (0.011)
MMDOS [21]	0.72 $\pm$ (0.0181)	0.78 $\pm$ (0.0159)	0.68 $\pm$ (0.0227)	0.69 $\pm$ (0.0193)	0.70 $\pm$ (0.0162)	0.76 $\pm$ (0.0122)	**0.84 $\pm$ (0.0256)**	0.061 $\pm$ (0.013)
BAR [8]	0.68 $\pm$ (0.0219)	0.71 $\pm$ (0.0192)	0.73 $\pm$ (0.0205)	0.70 $\pm$ (0.0125)	0.65 $\pm$ (0.0217)	0.69 $\pm$ (0.0155)	**0.82 $\pm$ ((0.0198)**	0.097 $\pm$ (0.028)

remove artifacts and extract temporal patterns. Labeling was done using majority voting within each window. We adopted a grid search fine-tuned window sizes and overlaps, ranging from 5% to 95% overlap and window durations from 0.1 to 1 second, to optimize parameter selection for capturing complex patterns across datasets.

2) Network Architecture and Training: Our experiments were conducted on a Linux server with an Intel i7-6850K CPU, 4x NVIDIA GeForce GTX 1080Ti GPUs, and 64 GB RAM. We used Python, specifically PyTorch, for data preprocessing and deep learning tasks. Given the class imbalance in most datasets, we adopted the macro-F1 score as the primary evaluation metric. The datasets were split into training, validation, and testing sets (60-20-20%), and hyperparameters were optimized using the validation set. Notably, the validation and test sets were not utilized during training.
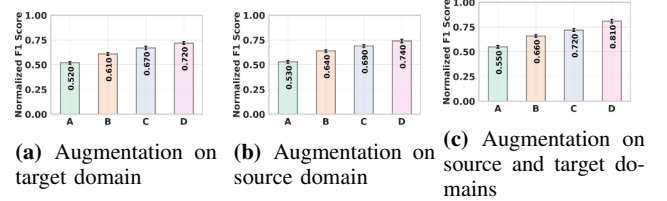
C. Results and Discussion

*1) **Classification Performance (RQ1):*** In this analysis, we compare and contrast μDAR against six state-of-the-art (SOTA) UDA algorithms, demonstrating a significant average improvement of $\approx$ **4-12%** across four public datasets. Our results distinguish the top-performing and second-best methods, revealing that SWL-Adapt frequently surpasses other UDA algorithms, likely due to its meta-optimized loss function, which assigns sample weights, enhances domain alignment, and improves cross-user adaptation, as shown in Table II.

Notably, μDAR excels across all datasets, covering activities from complex gym and sports-related actions to ADLs and iADLs. Its strength lies in using temporal ensembling to generate reliable pseudo-labels and integrating kernel-based class-wise conditional mean maximum discrepancy (kCMMD) to enhance domain adaptation. Additionally, performance improves with consistency regularization, effectively handling augmented sample variations in real-world settings and further supporting analyses are presented in section IV-C4.

(a) Feature Embedding on Target User using μDAR on the BAR dataset.

(b) Correlation between α parameter and uncertainty.

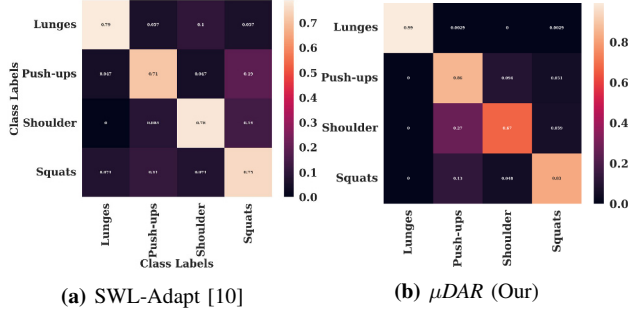**Fig. 2:** Ablation Analysis for **RQ1** and **RQ3**, respectively

(a) Augmentation on target domain

(b) Augmentation on source domain

(c) Augmentation on source and target domains

Fig. 3: The ablation study highlighting the performance difference between **(A)**, **(B)**, **(C)**, **(D)** corresponds to *Baseline*, *Temporal Ensembling*, *Temporal Ensembling + kCMMD loss* and *Temporal Ensembling + kCMMD loss + Consistency Regularizer*, respectively across all the benchmark datasets.

*2) **Robustness of μDAR to User-Variations (Proficiency levels) (RQ2):*** We assess the robustness and scalability of the μDAR framework by comparing across SOTA baseline algorithms for handling cross-user proficiency variations. The experiment trains models using expert data (source users) and evaluates their performance on beginner data (target users). Results show that μDAR significantly outperforms benchmarks, with an average macro F1 score improvement of **[7.3 $\pm$ (1.54)]%**. μDAR's success is due to its effective use of temporal ensembling, where the α parameter helps identify unique instance-based features across genders. By incorporating kCMMD loss and consistency, regularization captures non-linear features in high-dimensional space and ensures consistent predictions. Further analysis on optimizing the α parameter and its impact on pseudo-label quality is highlighted in section IV-C3.

We present the cross-proficiency user adaptation performance comparing the best accuracy SOTA SWL-Adapt algorithm and μDAR using the MM-DOS dataset and μDAR achieves an approximate **11%** improvement over SWL-Adapt as shown in Figure 4. Additionally, the results demonstrate that μDAR outperforms SOTA UDA algorithms in the wHAR domain across cross-user and cross-dexterity variations.

*3) **Compatibility Study (RQ3):*** In this study, we evaluated the effectiveness of self-ensembling for generating reliable pseudo-labels in the target domain. We used **entropy**, $\{H(x) = -\sum_{i=1}^{n} p(x_i) \log(p(x_i))\}$, to quantify prediction uncertainty. As shown in Figure 2b, there is an inverse relationship between the α parameter of ensembling and entropy: higher α leads to lower uncertainty, indicating more stable pseudo-labels. Our experiments found that an α range of *[0.55-0.75]* is optimal for generating high-quality pseudo-labels.

(a) SWL-Adapt [10]   **(b)** μDAR (Our)

Fig. 4: Confusion matrices for cross-dexterity levels (expert (source) $\rightarrow$ beginner (target)) on MMDOS dataset.

Figure 2b further validates that an α of **0.60** achieves faster convergence, emphasizing the critical role of α in improving pseudo-label quality for the target domain.

4) Comparative Analysis (RQ4): This analysis evaluates the individual and combined effects of temporal ensembling, kCMMD loss, and consistency regularization on μDAR's performance across four SOTA datasets. We report normalized F1 scores to demonstrate the synergy across three augmentation settings, as shown in Figure 3, highlighting the generalizability and robustness of μDAR. Our results show a significant improvement of **[22-26%]** when incorporating all methods over the baseline (no adaptation), as depicted in Fig. 3a, 3b, and 3c. This synergy enhances overall performance, advancing domain generalization for the wHAR domain.

V. CONCLUSION AND FUTURE WORK

This work introduces a novel single-stage joint optimization strategy for cross-user wHAR domain. 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. Additionally, using KL divergence for consistency regularization stabilizes label predictions in the presence of augmented samples. The μDAR framework demonstrated notable superiority over benchmark UDA algorithms in extensive experiments, showing generalizability and robustness, outperforming SOTA approaches by $\approx$ **4-12%**.

In the future, we plan to expand the capabilities of the μDAR 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.

REFERENCES

[1] Kerem Altun and Billur Barshan. Human activity recognition using inertial/magnetic sensor units. In *International workshop on human behavior understanding*, pages 38–51. Springer, 2010.

[2] Paulo Barbosa, Kemilly Dearo Garcia, João Mendes-Moreira, and André CPLF de Carvalho. 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*, pages 623–630. Springer, 2018.

[3] Chao Chen, Zhihang Fu, Zhihong Chen, Sheng Jin, Zhaowei Cheng, Xinyu Jin, and Xian-Sheng Hua. Homm: Higher-order moment matching for unsupervised domain adaptation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 3422–3429, 2020.

[4] Willian Dihanster Gomes de Oliveira and Lilian Berton. A systematic review for class-imbalance in semi-supervised learning. *Artificial Intelligence Review*, 56 (Suppl 2):2349–2382, 2023.

[5] Emadeldeen Eldele, Mohamed Ragab, Zhenghua Chen, Min Wu, Chee-Keong Kwoh, and Xiaoli Li. Cotmix: Contrastive domain adaptation for time-series via temporal mixup. *arXiv preprint arXiv:2212.01555*, 2022.

[6] Erik Englesson and Hossein Azizpour. Consistency regularization can improve robustness to label noise. *arXiv preprint arXiv:2110.01242*, 2021.

[7] Indrajeet Ghosh, Avijoy Chakma, Sreenivasan Ramasamy Ramamurthy, Nirmalya Roy, and Nicholas Waytowich. Permtl: A multi-task learning framework for skilled human performance assessment. In *2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 37–44. IEEE, 2022.

[8] Indrajeet Ghosh, Sreenivasan Ramasamy Ramamurthy, Avijoy Chakma, and Nirmalya Roy. Decoach: Deep learning-based coaching for badminton player assessment. *Pervasive and Mobile Computing*, page 101608, 2022.

[9] Henry Gouk, Eibe Frank, Bernhard Pfahringer, and Michael J Cree. Regularisation of neural networks by enforcing lipschitz continuity. *Machine Learning*, 110:393–416, 2021.

[10] Rong Hu, Ling Chen, Shenghuan Miao, and Xing Tang. 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*, volume 37, pages 6012–6020, 2023.

[11] Sara Khalifa, Guohao Lan, Mahbub Hassan, Aruna Seneviratne, and Sajal K Das. Harke: Human activity recognition from kinetic energy harvesting data in wearable devices. *IEEE Transactions on Mobile Computing*, 17(6):1353–1368, 2017.

[12] Samuli Laine and Timo Aila. Temporal ensembling for semi-supervised learning. *arXiv preprint arXiv:1610.02242*, 2016.

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

[14] Xiaofeng Liu, Chaehwa Yoo, Fangxu Xing, Hyejin Oh, Georges El Fakhri, Je-Won Kang, Jonghye Woo, et al. Deep unsupervised domain adaptation: A review of recent advances and perspectives. *APSIPA Transactions on Signal and Information Processing*, 11(1), 2022.

[15] M Jehanzeb Mirza, Jakub Micorek, Horst Possegger, and Horst Bischof. The norm must go on: Dynamic unsupervised domain adaptation by normalization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14765–14775, 2022.

[16] Sreenivasan Ramasamy Ramamurthy, Indrajeet Ghosh, Aryya Gangopadhyay, Elizabeth Galik, and Nirmalya Roy. Star: A scalable self-taught learning framework for older adults' activity recognition. In *2021 IEEE International Conference on Smart Computing (SMARTCOMP)*, pages 121–128. IEEE, 2021.

[17] Attila Reiss and Didier Stricker. Creating and benchmarking a new dataset for physical activity monitoring. In *Proceedings of the 5th International Conference on PErvasive Technologies Related to Assistive Environments*, pages 1–8, 2012.

[18] Chuan-Xian Ren, Pengfei Ge, Dao-Qing Dai, and Hong Yan. Learning kernel for conditional moment-matching discrepancy-based image classification. *IEEE transactions on cybernetics*, 51(4):2006–2018, 2019.

[19] Yong Ren, Jun Zhu, Jialian Li, and Yucen Luo. Conditional generative moment-matching networks. *Advances in Neural Information Processing Systems*, 29, 2016.

[20] Andrea Rosales Sanabria, Franco Zambonelli, Simon Dobson, and Juan Ye. Contrasgan: Unsupervised domain adaptation in human activity recognition via adversarial and contrastive learning. *Pervasive and Mobile Computing*, 78:101477, 2021.

[21] Ahmed Sharshar, Ahmed Fayez, Ahmed Abo Eitta, and Walid Gomaa. Mm-dos: A novel dataset of workout activities. In *2022 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2022.

[22] Maja Stikic, Kristof Van Laerhoven, and Bernt Schiele. Exploring semi-supervised and active learning for activity recognition. In *2008 12th IEEE International Symposium on Wearable Computers*, pages 81–88. IEEE, 2008.

[23] Xu Sun, Hisashi Kashima, Ryota Tomioka, Naonori Ueda, and Ping Li. A new multi-task learning method for personalized activity recognition. In *2011 IEEE 11th International Conference on Data Mining*, pages 1218–1223. IEEE, 2011.

[24] Remi Tachet des Combes, Han Zhao, Yu-Xiang Wang, and Geoffrey J Gordon. Domain adaptation with conditional distribution matching and generalized label shift. *Advances in Neural Information Processing Systems*, 33:19276–19289, 2020.

[25] Terry T Um, Franz MJ Pfister, Daniel Pichler, Satoshi Endo, Muriel Lang, Sandra Hirche, Urban Fietzek, and Dana Kulić. 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*, pages 216–220, 2017.

[26] Junqing Xie, Dong Wen, Lizhong Liang, Yuxi Jia, Li Gao, Jianbo Lei, 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):e9754, 2018.

[27] Qizhe Xie, Zihang Dai, Eduard Hovy, Thang Luong, and Quoc Le. Unsupervised data augmentation for consistency training. *Advances in neural information processing systems*, 33:6256–6268, 2020.

[28] Shuo Yang, Yijun Dong, Rachel Ward, Inderjit S Dhillon, Sujay Sanghavi, and Qi Lei. Sample efficiency of data augmentation consistency regularization. In *International Conference on Artificial Intelligence and Statistics*, pages 3825–3853. PMLR, 2023.