SFDA: Domain Adaptation With Source Subject Fusion Based on Multi-Source and Single-Target Fall Risk Assessment

Shibin Wu[©], Lin Shu[©], *Member, IEEE*, Zhen Song[©], and Xiangmin Xu[©], *Senior Member, IEEE*

Abstract— In cross-subject fall risk classification based on plantar pressure, a challenge is that data from different subjects have significant individual information. Thus, the models with insufficient generalization ability can't perform well on new subjects, which limits their application in daily life. To solve this problem, domain adaptation methods are applied to reduce the gap between source and target domain. However, these methods focus on the distribution of the source and the target domain, but ignore the potential correlation among multiple source subjects, which deteriorates domain adaptation performance. In this paper, we proposed a novel method named domain adaptation with subject fusion (SFDA) for fall risk assessment, greatly improving the cross-subject assessment ability. Specifically, SFDA synchronously carries out source target adaptation and multiple source subject fusion by domain adversarial module to reduce source-target gap and distribution distance within source subjects of same class. Consequently, target samples can learn more task-specific features from source subjects to improve the generalization ability. Experiment results show that SFDA achieved mean accuracy of 79.17 % and 73.66 % based on two backbones in a cross-subject classification manner, outperforming the state-of-the-art methods on continuous plantar pressure dataset. This study proves the effectiveness of SFDA and provides a novel tool for implementing cross-subject and few-gait fall risk assessment.

Index Terms— Cross-subject, domain adaptation, subject fusion, plantar pressure, fall risk assessment.

I. Introduction

IT IS commonly known that falls are a major public health issue, seriously affecting the life quality of the elderly due to their high frequency of onset, expensive cost of treatment, and longtime recovery [1], [2], [3]. Unlike fall detection [4], [5], fall risk assessment is an approach that focuses on early prevention and risk diagnosis through the development of sensor-based facilities [6]. In the field of sensor-based health monitoring, inertial sensors are widely used and have shown good accuracy in fall risk assessment [11], [12], [13], [14], [15], [16], [17]. However, when it comes to ease of use for the elderly, inertial sensor-based facilities typically require deployment on multiple parts of the body to achieve reliable performance [11], [12], [18], [19]. In contrast, smart shoes with integrated pressure sensor offer both comfort and practicality [7], [8], [66]. Consequently, assessing fall risk through few-gait plantar pressure monitoring is suitable for elderly in their daily lives [9], [10].

In classification of plantar pressure data, each subject can be considered as an independent domain because of their individual information, leading to domain discrepancies. When dealing with un-seen subjects, a well-trained model based on supervised learning struggles to perform well due to the domain discrepancies. Therefore, the objective of fall risk assessment is to address distribution mismatches and develop a model that can be generalized to well classify novel subjects, providing practical real-world applications and clinical diagnoses.

To better tackle cross-domain issues, unsupervised domain adaptation methods (UDA) are being explored to mitigate the domain shift between single-source and target distributions [20], [21], [23]. UDA methods can be categorized as two groups. The first one consists of discrepancy-based methods that aim to align source-target shifts by optimizing distribution distance using various measurement criteria, such as Maximum Mean Discrepancy (MMD), CORAL, KL, Maximum Density Divergence (MDD), within a common feature space

[24], [25], [26], [27], [28], [29], [30], [63]. The second one includes adversarial-based methods that apply adversarial modules to induce domain confusion and alleviate distributions difference [31], [32], [33], [34], [35].

More recently, another new setting called multi-source has emerged, and accordingly, multiple subjects must exist within the source domain in continuous plantar pressure classification and this could be defined as multi-source and single-target domain setting (MSST). Nevertheless, above single-source domain adaptation (SDA) approaches only focus on adapting source-target domain, regardless of marginal [25], conditional [62] or joint distribution adaptation [26], and ignore the potential correlation among sources [36], [37], [38], [39], [40]. Specifically, since domain shift also exists among different sources, SDA methods which mix multiple training subjects' samples as a combined source domain cannot guarantee that target samples will be effectively adapted to source subjects' distribution [36].

Similar to SDA, a straightforward approach for multi-source domain adaptation (MDA) to deal with multi-source data is also to merge all sources into one domain [39], which leads to an insufficient variance elimination in MSST [36]. In order to fully exploit multiple subjects' data distribution, some MDA began exploring feature representation approaches and combination of pre-learned classifiers [39], [40], [41], [42], [43]. The former approaches try to align the latent space of different domains based on optimizing the discrepancy loss, such as Rényi-divergence [48], L2 distance [49] or align the features through adversarial objectives, such as GAN loss [57], Wasserstein distance [58], [69]. The later approaches attempt to train per source separately and pairwise align the target with each source [37], [41], [48], [69]. Another solution is to assign a weight for each pre-learned classifier according to the relationship between source and target domain [39], [58].

However, many of these existing MDA approaches have notable limitations that the number of parameters and computational complexity in the model would sharply increase as more source subjects are involved when employing individual source classifiers or unshared feature extractors [37]. In our MSST setting, this strategy would require over 30 source classifiers, which is impractical. Even in a shared feature space, the scattering of multi-source representations degrades the effect of task-specific training (MSST fall risk assessment in this paper) for MDA [68]. Thus, a better solution is to alleviate source subjects' distribution shifts, aiming at fully exploiting task-specific features from more source subjects.

Consequently, it is necessary to reduce the domain gap among source domains from same class (referred as source subject fusion, SF) to avoid neglecting task-specific information of each source subject. SF, which is based on domain adversarial module, helps extracting task-specific features from as many source subjects as possible. To limit the complexity of model, a shared feature extractor is used to map data into a common space, and two discriminators are employed to reduce the domain discrepancy between disparate source subjects and alleviate the distance between source and target domain. Then, the generalization ability and robustness of the model can be enhanced by prompting SF and source-target domain adaptation (DA). Noting that in MSST, as the information from single-target subjects is unknown, we can regard them as a whole domain and we can interchangeably use the term "domain" and "subject" in this paper. Our main contributions are summarized as follows:

- We propose an end-to-end MDA framework by promoting SF and DA synchronously, where multiple source subjects are fused by adopting domain adversarial module to reduce their distribution distance.

- We illustrate the relation between SF and DA, that is, reducing the discrepancy within source subject encourages reducing the upper bound of the discrepancy between source domain and target domain, and the optimal weight of the SF and DA in training is verified by experiment.

- Empirically the proposed SFDA method can significantly outperform the state-of-the-art methods in continuous plantar pressure dataset.

This paper is organized as follows. In the Section I, the overview of fall risk assessment is summarized. In the Section II, the related work of fall risk classification, unsupervised domain adaptation and multi-source domain adaptation are investigated. In Section III, we illustrate the methodology, including problem statement, concrete modules of SFDA, training procedure and theoretical analysis of subject fusion. In Section IV, all experimental details and results are demonstrated and discussed. Finally, the summary based on this work are concluded in Section V.

II. RELATED WORK

A. Fall Risk Assessment

Due to the portability and wearable comfort, existing inertial datasets are not suitable for daily assessment, while plantar pressure could be used for long-term monitoring. Existing methods can be summarized as conventional machine learning and deep learning [6]. Through constructing manual bipedal or weak foot features, simple classifiers can output results at the cost of scalability [44], [45]. Dispense with feature engineering, deep learning methods could also output results in real-time by fewer raw gaits. Liang et al. [9] implemented ConvLSTM trained on raw plantar data. Nevertheless, non-cross-subject validation results in inflated accuracy (information leakage) [14]. Tunca et al. [11] compared LSTM and traditional classifiers based on spatio-temporal gait parameters in hold-out validation. Meyer et al. [12] evaluated the performance through LOSO validation. However, they overlook eliminating the domain shift between source and target data. With DG-DANN as the benchmark [46], Wu et al. [10] proposed a hierarchical framework to improve sensitivity and generalization ability. But generalization methods require comprehensive samples which cannot be guaranteed in real-world application.

B. UDA

For generalizing well when target samples are un-seen, many shallow layers approaches reduce domain gap according to various discrepancy criteria, such as MK-MMD, JDD [25], [26] while some statistic approaches achieve alignment in high dimensional space [24], [29]. Long et al. [63] proposed

a deep network for marginal distribution adaptation based on multi-layer and multi-kernel MMD. JAN was constructed in [26] for eliminating the joint distribution gap between source and target domain. Wang et al. [65] evaluated the weight of domain adaptation between marginal and conditional distribution. Deep adversarial approaches are another useful method for domain adaptation, which alleviate distribution difference by confusing the domain discriminator. Ganin et al. [32] proposed DANN with gradient reverser layer to realize adversarial function. Kumar et al. [34] proposed Co-DA to align global and local distribution at the same time. Fan et al. [70] adapted the source models on the target domain by modulating the domain-specific statistics in features of BN layers in sleep staging tasks.

C. MDA

Unlike SDA, it is necessary to make full use of every subject within training set [36], which is also important in fall risk assessment. Zhang et al. [47] aligned each source subject and the target data in individual feature spaces for sEMG classification. Wang et al. [68] proposed clustering embedded adversarial training with dependent source representations for multi-source sentiment analysis and digit classification. Zhu et al. [50] alleviated the conditional distribution in subdomain across different domains based on a local maximum mean discrepancy (LMMD) without adversarial training. Dai et al. [39] constructed extractors for each domain to separate shared and privacy features and assign different weight to different domain. Karimpour et al. [43] employed $l_{2,1}$ norm to re-weight the multiple source data for reducing the impact of unrelated source samples in image classification. Deng et al. [41] trained separated classifiers to align each source and target data in ECG classification together with sample-imbalance aware mixing strategy. Similarly, Liu et al. [71] trained DANN based on each source domain and the specified target domain. Wei et al. [72] aligned the data distribution for each pair of subjects and output by decision fusion.

III. METHODOLOGY

In this section, we will present the details on problem statement, concrete modules of SFDA, training procedure and theoretical analysis of subject fusion based on fall risk assessment.

A. Problem Statement

In the field of continuous plantar pressure classification, each participants' data space can be seen as an independent domain, which is defined as a joint distribution $\mathbb{P}^d(x, y)$ on $\mathcal{X} \times \mathcal{Y}$, where $\mathcal{X}, \mathcal{Y}$ are relatively input space and output space. In MSST setting, as the cost of obtaining source-domain subject labels is very low, $d \in \mathcal{D}\{1, \ldots, S\}$ denote the individual subjects within source domain. Accordingly, $\mathbb{P}^{\tilde{d}}(x)$ is defined as distribution from target domain, where $\tilde{d} \in \tilde{\mathcal{D}}\{1, \ldots, T\}$ and $\mathcal{D} \cap \tilde{\mathcal{D}} = \emptyset$. Here, the marginal probability distribution $\mathbb{P}^d(x)$ can be obtained, as well as the conditional distribution $\mathbb{P}^d(y|x)$, where $d \in \mathcal{D}$. In MSST, we assume that $\mathbb{P}^i(x) \neq \mathbb{P}^j(x)$ and $\mathbb{P}^i(y|x) \approx \mathbb{P}^j(y|x)$, $\forall i \neq j$ where

$i, j \in \mathcal{D}$ and there is no need to concern about the ill-posed problem for each subject was labelled as one class, which can hold the setting that the conditional distributions stay stably.

Specifically, $\{(\mathbf{X}^d, \mathbf{y}^d) \sim \mathbb{P}^d(x, y)\}_{d=1}^S$ are defined as training data and each $\mathbf{X}^d$ contains N samples $\mathbf{x}_m^d$ where $m \in \{1, \ldots, N\}$. Note that $y_m^d \in \{0, 1\}$, $\forall m \in \{1, \ldots, N\}$ where $\{0\}$ represents "low risk" and $\{1\}$ represents "high risk" and $u_m^d \in \{1, \ldots, S\}$ indicates subject label, which could also be understood as subject ID in source domain. Un-seen data from T target domains $\{(\mathbf{X}^{\tilde{d}})\}_{\tilde{d}=1}^T$ are defined as testing data. Data space with domain label $\mathcal{U} = \{(\mathbf{X}^d, 0)\}_{d=1}^S \cup \{(\mathbf{X}^{\tilde{d}}, 1)\}_{\tilde{d}=1}^T$ is constructed, where $d \in \mathcal{D}$, $\tilde{d} \in \tilde{\mathcal{D}}$ and the samples from S and T are labeled $o_m^d = 0$ and $o_n^{\tilde{d}} = 1$, respectively, where n indicates the number of testing samples. Based on S, our goal is to train a model $f : \mathcal{X} \rightarrow \mathcal{Y}$ that can assess fall risk correctly when facing un-seen subject samples $\mathbf{X}^{\tilde{d}}$ whose distributions are unknown. Compared with traditional fall risk assessment protocols, MSST avoids any information leakage from target domain, which is more in line with practical application.

B. Source Subject Fusion

In MSST DA, especially in plantar pressure classification, neither a relevant feature from each source subject should be ignored or priorly downgraded. And it is worth noticing that, besides dataset shift [51], the shift between subjects might include more recessive reasons, such as their waking habit, weight, mental health, disease history, the devices or time that collected data [10], [14], which reflect in the discrepancy in source subjects. Thus, target samples are difficult to transfer to all the source domains in the same class due to above inevitable characteristics. However, SDA might mix source subjects together and MDA require too many individual feature extractors. Consequently, we need target samples to be transferred to more source subjects in the same class to extract task-specific features, improving the generalization ability of model and interpretability in clinical diagnosis.

To illustrate the relationship between internal SF and external overall DA in statistic way, let $\mathcal{H}$ be Reproducing Kernel Hilbert Spaces (RKHS), and $\phi : \mathcal{X} \rightarrow \mathcal{H}$, where ϕ is a representation function that maps the instance set to feature space. For simplicity, we assume that there are only two subjects in the source domain and MMD is used to measure the distance or discrepancy between two distributions which are mapped into $\mathcal{H}$ [53]. The empirical estimate of the distance between $\mathbf{X} \sim \mathbb{P}(x)$ and $\mathbf{Y} \sim \mathbb{P}(y)$, as defined by MMD, is

$$\text{Dist}(\mathbf{X}, \mathbf{Y}) = \left\| \frac{1}{m1} \sum_{i=1}^{m1} \phi(x_i) - \frac{1}{m2} \sum_{j=1}^{m2} \phi(y_j) \right\|_{\mathcal{H}} \quad (1)$$

where $\mathbf{X} = \{x_1, \ldots, x_{m1}\}$ and $\mathbf{Y} = \{y_1, \ldots, y_{m2}\}$ are random variable sets. Thus, considering the independent domain distribution in $\mathcal{H}$, two kinds of distance can be well-estimated by MMD: 1) distance between source and target domain, 2) distance between each subject inside source domain.

Assuming that $\mathbf{X}^1 \sim \mathbb{P}^1(x)$, $\mathbf{X}^2 \sim \mathbb{P}^2(x)$ in source domain $\mathcal{D}$, where $\mathbb{P}^1(x) \neq \mathbb{P}^2(x)$. And $\mathbf{X}^{\tilde{d}}$ is defined as samples from

target domain, where $\tilde{d} \in \tilde{\mathcal{D}}$ and $\mathbf{X}^{\tilde{d}} = \{\tilde{x}_1, \ldots, \tilde{x}_{\text{m}\tilde{d}}\}$. The distance between source and target domain can be approximately estimated as the sum of the distances between two subjects and $\tilde{\mathcal{D}}$:

$$\text{Dist}(\mathcal{D}, \tilde{\mathcal{D}}) \triangleq \text{Dist}(\mathbf{X}^1, \mathbf{X}^{\tilde{d}}) + \text{Dist}(\mathbf{X}^2, \mathbf{X}^{\tilde{d}}) \qquad (2)$$

As shown in (1) and (2), $\text{Dist}(\mathcal{D}, \tilde{\mathcal{D}})$ and the distance between the above two subjects can be respectively empirically calculated by the squared MMD and written as:

$$
\begin{aligned}
&\text{Dist}(\mathcal{D}, \tilde{\mathcal{D}}) \\
&= \text{tr}\left(\mathbf{C}^T \left[\mathbf{X}^1 \ \mathbf{X}^{\tilde{d}}\right] \begin{bmatrix} \frac{1}{\text{m1}^2}\mathbf{11}^T & \frac{-1}{\text{m1m}\tilde{d}}\mathbf{11}^T \\ \frac{-1}{\text{m1m}\tilde{d}}\mathbf{11}^T & \frac{1}{\text{m}\tilde{d}^2}\mathbf{11}^T \end{bmatrix} \begin{bmatrix} \mathbf{X}^1 \\ \mathbf{X}^{\tilde{d}} \end{bmatrix} \mathbf{C}\right) \\
&+ \text{tr}\left(\mathbf{C}^T \left[\mathbf{X}^2 \ \mathbf{X}^{\tilde{d}}\right] \begin{bmatrix} \frac{1}{\text{m2}^2}\mathbf{11}^T & \frac{-1}{\text{m2m}\tilde{d}}\mathbf{11}^T \\ \frac{-1}{\text{m2m}\tilde{d}}\mathbf{11}^T & \frac{1}{\text{m}\tilde{d}^2}\mathbf{11}^T \end{bmatrix} \begin{bmatrix} \mathbf{X}^2 \\ \mathbf{X}^{\tilde{d}} \end{bmatrix} \mathbf{C}\right)
\end{aligned}
\qquad (3)
$$

$$
\begin{aligned}
&\text{Dist}(\mathbf{X}^1, \mathbf{X}^2) \\
&= \left\| \frac{1}{\text{m1}} \sum_{i=1}^{\text{m1}} \phi(x_i) - \frac{1}{\text{m2}} \sum_{j=1}^{\text{m2}} \phi(x_j) \right\|_{\mathcal{H}}^2
\end{aligned}
\qquad (4)
$$

where $\mathbf{C}$ is transformation matrix. It is known that there must be a subject in source domain is closer to $\tilde{\mathcal{D}}$ in the $\mathcal{H}$ space. Assume that $\mathbf{X}^1$ is closer to $\tilde{\mathcal{D}}$ than $\mathbf{X}^2$ in a $\mathcal{H}$ and we have:

$$\text{Dist}(\mathcal{D}, \tilde{\mathcal{D}}) \leqslant 2 \times \text{Dist}(\mathbf{X}^2, \mathbf{X}^{\tilde{d}}) \qquad (5)$$

Actually, the gap between different subjects will be increased when data are collected after long time interval or with different devices under the assumption that $\mathbf{X}^1$, $\mathbf{X}^2$, $\mathbf{X}^{\tilde{d}}$ were subjects who have individual information without significant differences, such as their ages, genders, heights, weights, health conditions. Therefore, the distribution discrepancy between source samples collected in the early stage and target samples to be transferred is likely to be greater than that between source subjects. Thus, we have:

$$\text{Dist}(\mathbf{X}^2, \mathbf{X}^{\tilde{d}}) \geqslant \text{Dist}(\mathbf{X}^1, \mathbf{X}^2) \qquad (6)$$

The detailed reason for assuming equation (6) is listed in Appendix. According to (3), (4), (5), (6) and assuming that the representation distribution of $\mathbf{X}^1$ stay stably in $\mathcal{H}$, we can learn that, on a finite probability space, reducing the proposed discrepancy between subjects in source domain could push $\text{Dist}(\mathbf{X}^2, \mathbf{X}^{\tilde{d}})$ be closer to $\text{Dist}(\mathbf{X}^1, \mathbf{X}^{\tilde{d}})$. Naturally, the upper bound of discrepancy between source and target domain could be reduced since $\text{Dist}(\mathbf{X}^2, \mathbf{X}^{\tilde{d}})$ becomes smaller under subject fusion.

In detail, besides classification task, another purpose of domain adaptation with subject fusion is to find a transform that can reduce the distance between $\mathcal{D}$ and $\tilde{\mathcal{D}}$, as well as confuse the distribution of subjects in $\mathcal{D}$, which can be optimized as:

$$
\begin{aligned}
&\min_{\mathbf{C}} \ \text{Dist}(\mathcal{D}, \tilde{\mathcal{D}}) + \lambda \|\mathbf{C}\|_F \\
&\min_{\mathbf{C}} \ \text{Dist}(\mathbf{X}^1, \mathbf{X}^2) + \gamma \|\mathbf{C}\|_F
\end{aligned}
\qquad (7)
$$

where λ, γ are the tradeoff parameters for transformation penalty. In other words, when minimize the distance between source domain and target domain, we require that the discrepancy between source subjects will not be too large, aiming at adapting more efficiently.

C. Model Summary

According to the concept of subject fusion, we could easily transfer the two-subject setting, as shown in equation (7), to multi-subject setting. In this part, we will concretely introduce the modules in SFDA framework, as shown in Fig. 1. Here, binary classifier L, D and multi-class classifier F all utilize cross entropy losses for optimizing the gradients, which is denoted as $\mathcal{L}_L$, $\mathcal{L}_D$, $\mathcal{L}_F$, respectively. Given a shared feature extractor $E(\cdot; \theta_e)$ with parameters θ_e, a task classifier $L(\cdot; \theta_l)$ with parameters θ_l, a domain discriminator $D(\cdot; \theta_d)$ with parameters θ_d, a subject fusion discriminator $F(\cdot; \theta_f)$ with parameters θ_f, the goal of SFDA could be achieved by the following modules.

1) The Shared Feature Extractor E: Like RKHS, the shared feature extractor E aims at producing a common space that could allow features transformation $\mathbf{C}$ and extract deep shared features from both source and target domain. The tradeoff parameters λ, γ are achieved by the L2 regularization in all experiments.

2) The Label Classifier L: L is a usual classifier and its input $E(\mathbf{x}_m^d)$ all come from source domain. The formulated objective of L is:

$$\underset{\theta_e, \theta_l}{\arg\min} \ \frac{1}{N \cdot S} \sum_{d=1}^{S} \sum_{\text{m}=1}^{N} E_{x \sim X^d} \left[\mathcal{L}_L(L(E(\mathbf{x}_m^d)), \mathbf{y}_m^d) \right] \qquad (8)$$

where $\mathbf{y}_m^d$ refer to the class label of source sample $\mathbf{x}_m^d$, $\forall m \in \{1, \ldots, N\}$ and d refer to the subject label within source domain, $\forall d \in \mathcal{D} = \{1, \ldots, S\}$. As equation (8) demonstrated, this optimization pushes L to access fall risk correctly among labelled source data.

3) The Subject Fusion Discriminator F: Here, the second optimization objective of equation (7) could be achieved by F. It is complex and time-consuming to calculate the MMD between each pair of samples between two different subjects within source domain. Hence, motivated by multi-source training methods [39], [46], we adopt adversarial module to achieve the goal that alleviating the domain gap between source subjects. It is obvious that MMD tends to be 0 when the distributions become similar or confused, thus the MMD optimization could be replaced by distribution fusion. The formulated objective of F is:

$$\underset{\theta_e, \theta_f}{\arg\min} \ \frac{1}{N \cdot S} \sum_{d=1}^{S} \sum_{\text{m}=1}^{N} E_{\mathbf{x} \sim \mathbf{X}^d} \left[\mathcal{L}_F(G(F(E(\mathbf{x}_m^d))), \mathbf{u}_m^d) \right] \quad (9)$$

where $\mathbf{u}_m^d$ is the subject ID label for source sample $\mathbf{x}_m^d$, $\forall m \in \{1, \ldots, N\}$. $G(\cdot)$ indicates Gradient Reverse Layer (GRL) that could reverse the gradient in backpropagation. In the training procedure when $G(\cdot)$ is frozen, the goal
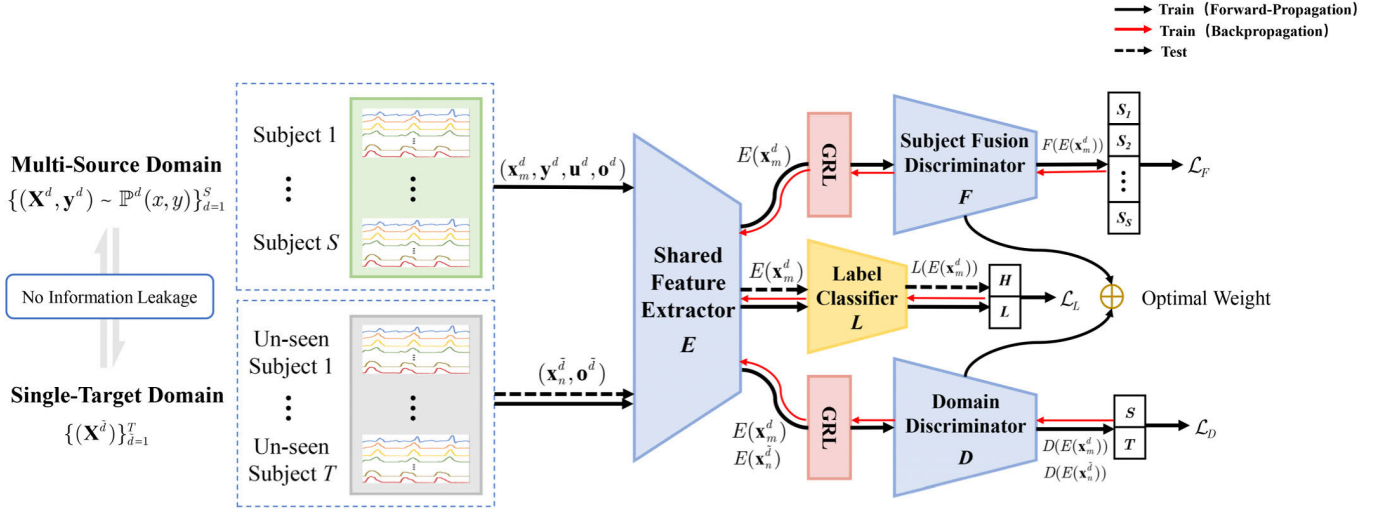
Fig. 1. The framework of the proposed SFDA model. The shared feature extractor E captures features $E(\mathbf{x}_m^d)$ and $E(\mathbf{x}_n^{\tilde{d}})$ from labeled source data $\mathbf{x}_m^d$ with {class label: $\mathbf{y}^d$; subject label: $\mathbf{u}^d$; domain label: $\mathbf{o}^d$} and unlabeled target data $\mathbf{x}_n^{\tilde{d}}$ with {domain label: $\mathbf{o}^{\tilde{d}}$}. First, $E(\mathbf{x}_m^d)$ with $\mathbf{y}^d$ is fed into label classifier to extract task-specific features. Next, $E(\mathbf{x}_m^d)$ with $\mathbf{u}^d$ is fed into subject fusion discriminator F to alleviate the internal subject shift within source domain through Gradient Reverse Layer (GRL) in backpropagation. Then, $E(\mathbf{x}_m^d)$ and $E(\mathbf{x}_n^{\tilde{d}})$ along with their domain label are synchronously fed into domain discriminator D to reduce the distribution shift between source and target data. In order to figure out the interaction between SF and DA, different weight biases are adopted in $\mathcal{L}_F$ and $\mathcal{L}_D$. According to the optimal weight, the parameters in F and D could be updated, which results in the well-generalized model.

Algorithm 1 Training on **SFDA**

Require: labeled source data $\{\mathbf{x}^d, \mathbf{y}^d\}$; unlabeled target data $\mathbf{x}^{\tilde{d}}$; source subjects label $\mathbf{u}^d$; domain label $\{\mathbf{o}_m^d, \mathbf{o}_n^{\tilde{d}}\}$; Hyperparameter $\alpha \in [0, 1]$, b, $t \in \mathbb{N}$
1: Initialize parameters of E, L, F, D
2: **repeat**
3:     **for** $i = 1$ **to** t **do**
4:         $\mathcal{L}$, $\mathcal{L}_L$, $\mathcal{L}_D$, $\mathcal{L}_F = 0$
5:         Sample a mini-batch $\mathbf{x} \triangleq (\mathbf{x}_m^d, \mathbf{u}_m^d, \mathbf{y}_m^d, \mathbf{o}_m^d)_{m=1}^b$
6:         Sample a mini-batch $\tilde{\mathbf{x}} \triangleq (\mathbf{x}_n^{\tilde{d}}, \mathbf{o}_n^{\tilde{d}})_{n=1}^b$
7:         Feed $\mathbf{x}$, $\tilde{\mathbf{x}}$ into E and get $E(\mathbf{x})$, $E(\tilde{\mathbf{x}})$
8:         Feed $E(\mathbf{x})$ into L and get $L(E(\mathbf{x}))$
9:         Feed $E(\mathbf{x})$ into F and get $F(E(\mathbf{x}))$
10:       Feed $E(\mathbf{x})$, $E(\tilde{\mathbf{x}})$ into D and get
               $D(E(\mathbf{x}))$, $D(E(\tilde{\mathbf{x}}))$
11:       $\mathcal{L}_L += \mathcal{L}_L(L(E(\mathbf{x})), \mathbf{y}^d)$
12:       $\mathcal{L}_F += \mathcal{L}_F(F(E(\mathbf{x})), \mathbf{u}^d)$
13:       $\mathcal{L}_D += \mathcal{L}_D(D(E(\mathbf{x})), \mathbf{o}^d) + \mathcal{L}_D(D(E(\tilde{\mathbf{x}})), \mathbf{o}^{\tilde{d}})$
14:       $\mathcal{L} = \mathcal{L}_L + \mathcal{L}_F + \alpha \cdot \mathcal{L}_D$
15:       Updating E, L, F, D to minimize $\mathcal{L}$
16: **until** convergence

---

of F is to distinguish exactly where the source samples come from. After gradient reverse through GRL, as equation (9) presents, F makes the parameters in E hard to extract source domain-specific features, which achieves the goal that alleviates the subject distribution shift within source domain.

4) The Domain Discriminator D : Here, the first optimization objective of equation (7) could be achieved by D. Similar to F, D is a usual adversarial module (binary classifier) that

pays attention on source and target domain and its formulated objective is:

$$\arg\min_{\theta_e, \theta_d} \frac{1}{N \cdot S} \sum_{d=1}^{S} \sum_{m=1}^{N} E_{\mathbf{x} \sim \mathbf{X}^d} \left[\mathcal{L}_D(G(D(E(\mathbf{x}_m^d))), \mathbf{o}_m^d) \right]$$
$$+ \frac{1}{\tilde{N} \cdot T} \sum_{\tilde{d}=1}^{T} \sum_{n=1}^{\tilde{N}} E_{\mathbf{x} \sim \mathbf{X}^{\tilde{d}}} \left[\mathcal{L}_D(G(D(E(\mathbf{x}_n^{\tilde{d}}))), \mathbf{o}_n^{\tilde{d}}) \right] \quad (10)$$

where $\mathbf{o}_m^d$, $\mathbf{o}_n^{\tilde{d}}$ refer to the source and target domain label and $\tilde{d} \in \tilde{\mathcal{D}} = \{1, \ldots, T\}$. As shown in equation (10), in the training procedure when $G(\cdot)$ is frozen, the binary classifier D plans to precisely identify which domains do the samples come from. When updating parameters of model in backpropagation, this optimization is to align distribution from source and target domain through GRL. When the training of D is converged, distribution gap between source and target data could be reduced. In other words, MMD between source and target domain could also be alleviated without complex computing. Noting that GRL will not be used in training forward stage.

5) The Training Procedure: In general, the proposed end-to-end framework requires backpropagation to update the parameters in four modules, as demonstrated in Algorithm 1. Here, the realization of proposed optimization problem is summarized. Source data and target data will be input to E (could be seen as $\mathbf{C}$ in equation (7)) at the same time, while the features of source data will enter L, F, D and features of target data will only enter D. Concretely, as demonstrated in equation (7), optimizing $\mathcal{L}_D$ help minimizing $\text{Dist}(\mathcal{D}, \tilde{\mathcal{D}})$ and $\mathcal{L}_F$ help minimizing $\text{Dist}(\mathbf{X}^1, \mathbf{X}^2)$. Thus, we could satisfy the demand in optimization (7) by synchronously carrying out DA and SF. The total loss function of SFDA is formulated in

weighted summation formula:

$$\mathcal{L} = \mathcal{L}_L + \mathcal{L}_D + \alpha \cdot \mathcal{L}_F \qquad (11)$$

where α is non-training configuration that represent the loss weight of SF and can regulate the relation between SF and DA.

D. Theoretical Analysis

Aiming at better analyzing the generalization ability of our proposed SFDA framework, we report a universal theoretical generalization bound analysis in field of DA based on the theories from [54], [55], and [56].

Theorem 1: Let $\mathcal{H}$ be a hypothesis space and $\mathcal{S}$, $\mathcal{T}$ is referred as source domain, target domain, respectively. If we denote $h \in \mathcal{H}$ is a function that need to be tested in un-seen target domain, then for any $h \in \mathcal{H}$, we have:

$$R_{\mathcal{T}}(h) \leqslant R_{\mathcal{S}}(h) + d_{\mathcal{H}}(\mathcal{S}, \mathcal{T}) + C \qquad (12)$$

where $d_{\mathcal{H}}(\mathcal{S}, \mathcal{T})$ indicates the domain discrepancy between source domain $\mathcal{S}$ and target domain $\mathcal{T}$ and C can be seen as the shared error of ideal function $h^* \in \mathcal{H}$ for $\mathcal{S}$, $\mathcal{T}$. In addition, $R_{\mathcal{S}}(h)$, $R_{\mathcal{T}}(h)$ refer to the expected error on $\mathcal{S}$, $\mathcal{T}$, respectively. In fall risk assessment through SFDA in MSST setting, we can analyze the generalization error according to (12). Focusing on $d_{\mathcal{H}}(\mathcal{S}, \mathcal{T})$ while C could be disregarded as [25] and [32] did, we could assume that source domain involves S subjects and it is natural to extend (5) into multi-subject situation:

$$d_{\mathcal{H}}(\mathcal{S}, \mathcal{T}) \triangleq \sum_{i=1}^{S} d_{\mathcal{H}}(\mathbb{P}_i, \mathbb{P}_{\tilde{d}}) \leqslant S \cdot d_{\mathcal{H}}(\mathbb{P}_f, \mathbb{P}_{\tilde{d}}) \qquad (13)$$

where $\mathbb{P}_i$ indicates the i-th global distribution in $\mathcal{S}$, $\mathbb{P}_{\tilde{d}}$ indicates the global distribution of $\mathcal{T}$, $\mathbb{P}_f$ and $\mathbb{P}_{f*}$ denotes the distribution of source subject which is the most different and most similar with $\mathbb{P}_{\tilde{d}}$, respectively. Before training started, the largest distribution gap within multiple source subjects could be defined as:

$$d_{\mathcal{H}}(\mathbb{P}_f, \mathbb{P}_{f*}) > 0 \qquad (14)$$

According to equation (7) and (11), SFDA can achieve the goal of minimizing Dist($\mathbf{X}^1$, $\mathbf{X}^2$) (the distance within source subjects) and Dist($\mathcal{D}$, $\tilde{\mathcal{D}}$) (the distance between source and target domain) at the same time. When paying attention on optimizing Dist($\mathbf{X}^1$, $\mathbf{X}^2$) itself, parameters in SFDA modules can map the input data into a feature space that can narrow the distribution gap between disparate source subjects. The source subject distance can be denoted as:

$$\sum_{i=1}^{S} \sum_{j=1}^{S} d_{\mathcal{H}}(\mathbb{P}_i, \mathbb{P}_j), \forall i \neq j \qquad (15)$$

Herein, the result of equation (15) will be reduced when subject fusion training converges. Ideally, each subject in source domain will mix with each other and every personal distribution $\mathbb{P}_i$ can be regarded as $\mathbb{P}_{f*}$ in source domain. Meanwhile, when optimizing Dist($\mathcal{D}$, $\tilde{\mathcal{D}}$) (source-target domain adaptation), the distance between source subjects and target subjects will be decreased. Thus $d_{\mathcal{H}}(\mathbb{P}_{f*}, \mathbb{P}_{\tilde{d}})$ is reduced.

TABLE I
DOMAIN DIVISION IN CONTINUOUS PLANTAR PRESSURE DATASET

Division		Subjects
P1	High risk	#17, #21, #23, #26, #30, #46
	Low risk	#1, #2, #4, #5, #36, #41
P2	High risk	#18, #20, #27, #33, #49
	Low risk	#3, #7, #22, #29, #43, #50
P3	High risk	#12, #24, #28, #37, #39, #48
	Low risk	#6, #8, #9, #14, #42, #44, #45
P4	High risk	#19, #25, #31, #35, #40, #47
	Low risk	#10, #11, #13, #15, #32, #34

*P denotes the domain and # denotes the subject ID.

At this time, $d_{\mathcal{H}}(\mathcal{S}, \mathcal{T}) = S \cdot d_{\mathcal{H}}(\mathbb{P}_{f*}, \mathbb{P}_{\tilde{d}})$ and $R_{\mathcal{T}}(h)$ will be bounded. In conclusion, the synchronization of SF and DA can guarantee the generalization performance of the model.

IV. EXPERIMENT AND DISCUSSION

In this section, we introduce the dataset and experimental details for fall risk assessment such that an expert should be able to reproduce the main results. When discussing the assessing ability of SFDA, we empirically evaluate and compare the performance of proposed methods, including accuracy, sensitivity, boxplot results, proxy $-\mathcal{A}-$ distance and t-SNE visualization.

A. Experimental Setup

1) Dataset: As the same in our previous work [10], [45], we train and evaluate the proposed model in the continuous plantar pressure dataset, which can be used for fall risk assessment in long-term monitoring comfortably [67]. This plantar pressure dataset contains 48 subjects and 7462 samples in total, including 23 high risk subjects and 25 low risk subjects. As shown in Table I, the dataset is divided into four different MSST setting in order to traverse the whole dataset and produce comprehensive experimental results, as 4-fold cross validation did [79], [80]. Noting there is no information leakage between divisions when adopting leave-one-division-out.

2) Baselines: We compare proposed SFDA with the closely related state-of-the-art baselines. When compared with unsupervised single-source domain methods, all the subjects within source domain will be regarded as a whole single domain while neglect their internal discrepancy. All baselines in the comparison involve: **DAN [63]**, which reduces global discrepancy through MMD; **Deep CORAL [27]**, which reduces global distance through second-order statistics; **DANN [32]**, which alleviates global distance through adversarial module; **CDAN [62]**: which alleviates discrepancy through conditional adversary; **DSAN [50]**, which reduces global and local discrepancy in multi-subdomain adaptation; **MhNet [10]**, which construct a hierarchical network for fall risk assessment.

TABLE II
ACCURACY (%) OF DIFFERENT METHODS ON FOUR MSST SETTING BASED ON RESNET-50

Source	Target	ResNet-50	DAN	DeepCoral	DANN	CDAN	DSAN	MhNet	SFDA
P2, P3, P4	P1	77.11±1.0	80.28±0.9	84.27±0.3	86.85±0.3	84.38±0.0	80.92±0.8	<u>86.87±5.6</u>	**91.14±0.4**
P1, P3, P4	P2	56.39±0.3	58.46±0.0	58.28±0.9	60.27±1.4	56.93±0.0	64.22±0.4	**91.20±6.9**	<u>65.10±0.7</u>
P1, P2, P4	P3	73.20±1.0	76.62±0.8	72.69±0.0	76.85±0.8	<u>78.69±0.5</u>	76.52±0.5	46.00±0.7	**78.78±0.7**
P1, P2, P3	P4	70.37±0.8	77.05±0.4	72.88±0.2	73.85±0.6	<u>78.38±0.0</u>	77.02±0.9	66.95±2.4	**81.66±0.6**
Average		69.27	73.10	72.03	74.45	74.60	<u>74.67</u>	72.76	**79.17**

*The best performance is highlighted in bold, and the second-best performance is underlined.

TABLE III
ACCURACY (%) OF DIFFERENT METHODS ON FOUR MSST SETTING BASED ON VGGNET-16

Source	Target	VggNet-16	DAN	DeepCoral	DANN	CDAN	DSAN	MhNet	SFDA
P2, P3, P4	P1	68.78±0.9	74.34±0.0	69.46±1.4	70.57±1.0	71.26±0.0	77.05±0.3	**86.87±5.6**	<u>77.08±0.5</u>
P1, P3, P4	P2	54.79±1.2	61.46±0.1	58.09±0.9	55.52±0.7	57.07±0.0	56.13±0.1	**91.20±6.9**	<u>63.79±0.9</u>
P1, P2, P4	P3	63.77±0.6	71.24±0.5	71.49±0.0	69.65±0.9	74.56±0.2	<u>76.15±0.0</u>	46.00±0.7	75.38±0.9
P1, P2, P3	P4	72.44±0.4	75.37±0.1	75.82±1.3	76.28±0.5	71.44±0.0	<u>76.94±0.0</u>	66.95±2.4	**78.38±1.0**
Average		64.95	70.60	68.72	68.00	68.58	71.57	<u>72.76</u>	**73.66**

*The best performance is highlighted in bold, and the second-best performance is underlined.

3) Implementation Details:

3) Implementation Details: Above methods are reproduced in Pytorch and GeForce GTX 1080 Ti / GeForce RTX 2080 Ti. In pre-processing stage, raw data are split with overlapping of 2 gaits in time sequence. Then, 16-channel plantar pressure data are arranged vertically in sequence to form 3-gait samples, which are 2D tensors of 16×69 (channel × time points) as [10] did. In four MSST settings, each division Pi, $i \in \{1, 2, 3, 4\}$ refers to a multi-subject domain, which includes m subjects (#). When evaluating the performance of methods, each Pi will be taken in turn as a target dataset while the rest divisions will be taken as source training dataset. When testing Pi, $i \in \{1, 2\}$, the SGD optimizers select 10^{-3}, 10^{-2}, 10^{-2} as learning rate of pre-train model finetuning, classifier and adversarial module while 10^{-4}, 10^{-3}, 10^{-3} are used for Pi, $i \in \{3, 4\}$. α could be selected from [0, 1].

For our framework, ResNet-50 [59] and VggNet-16 [60] are adopted as the backbone network. A 2-layer fully-connected layer is used as classifier while adversarial module is set as 2 layers with 2048 nodes if needed. The batch size is set to 12 in ResNet-50 and 32 in VggNet-16, and the maximum training epoch is set to 100. The weight decay is fine-tuned from $\{5 \times 10^{-4}, 1 \times 10^{-3}, 2 \times 10^{-3}\}$ and momentum is set to 0.95. We choose the best performance under different configurations for each method. Specifically, Coral, MMD, LMMD loss weight are set as 0.5, 0.35, 0.2 for better convergence. Since MhNet is superior owing to hierarchical structure and voting mechanism without pre-trained model, it is fairer to discuss the average voting classification accuracy (Segment=7). Also, as MhNet did not use any domain adaptation method, it is not suitable to analyze the Proxy$-\mathcal{A}-$ distance or t-SNE visualization.

B. Experimental Results and Analysis

The cross-subject classification results are shown in Table II and Table III. Additional experiment results from Inception-

V3 [78], ResNet-34 [59] and VggNet-19 [60] are presented in Appendix. Each baseline is implemented in five replications, and SFDA outperforms all the compared baselines in terms of the average accuracy and MhNet votes and averages segment under five thresholds (10%, 20%, 30%, 40% and 50%).

When using ResNet-50 as pre-trained backbone, SFDA ranks first in 3 out of 4 MSST settings in Table II, achieving average accuracy of 79.17%. Particularly, SFDA outperforms the second-best baseline by nearly 3.3% in terms of accuracy when P4 is considered the target domain. In addition, SFDA demonstrates stable performance in all four MSST settings, while the performances of other baselines vary significantly across scenarios, highlighting the robustness of SFDA. Concretely, DeepCoral performs well when P1 is target domain but exhibits inferior accuracy in other settings, and even shows negative transfer when P3 is the target domain. MhNet achieves good performance when P1 and P2 are target domains but performs the worst when P3 is target domain. Especially in P2, MhNet outperform other methods by around 30%, which indicates the superiority of hierarchical structure and voting mechanism in certain fall risk assessment setting. In general, DSAN, which pays attention on multiple subdomains adaptation, performs relatively good in MSST settings. Both joint distribution adaptation and adversarial approaches are superior to methods that solely focus on marginal distribution alignment. When using VggNet-16 as pre-trained backbone, SFDA still performs stable and good in 4 MSST settings, as shown in Table III. MhNet performs well except for P3 and P4, which may result from the insufficient domain-variant feature extraction. Although SFDA achieves the best average accuracy, the overall performance is inferior to that of ResNet-50. This may be attributed to the shallower network's inability to extract enough task-specific features. Hence, compared to the baselines, SFDA significantly improves generalization
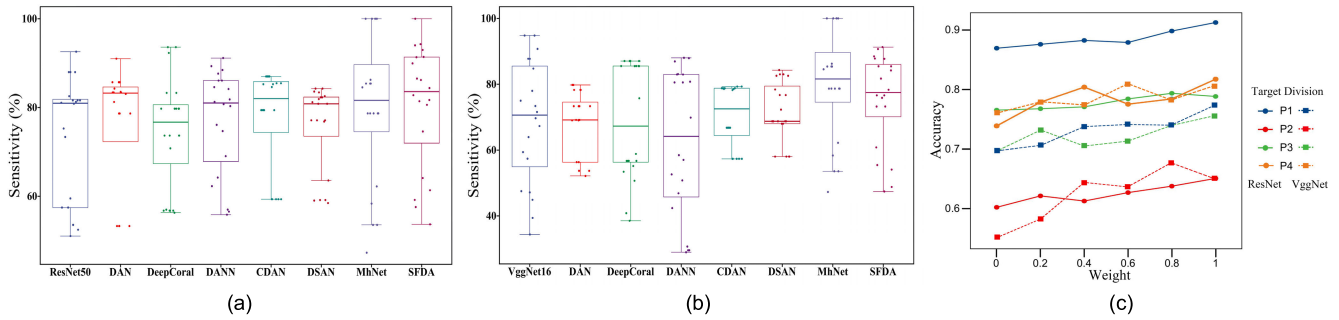
Fig. 2. (a) Average sensitivity of different methods in 4 MSST settings based on ResNet-50; (b) Average sensitivity of different methods in 4 MSST settings based on VggNet-16; (c) Ablation study on SF in 4 MSST settings.
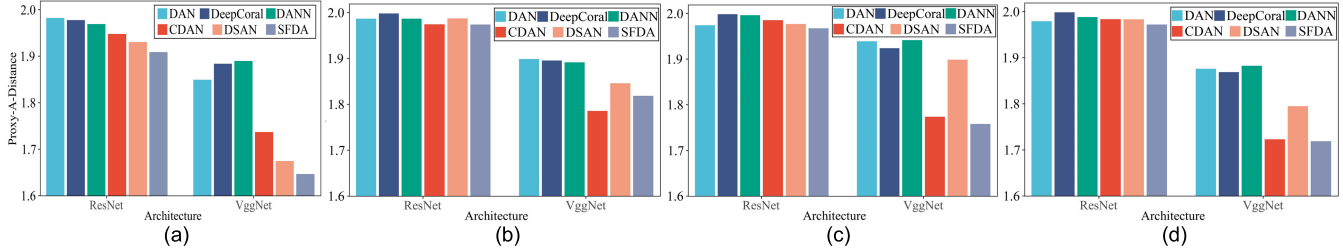

Fig. 3. Proxy$-\mathcal{A}-$ distance of different methods in MSST setting: (a) P2, P3, P4$\rightarrow$ P1; (b) P1, P3, P4$\rightarrow$ P2; (c) P1, P2, P4$\rightarrow$ P3; (d) P1, P2, P3$\rightarrow$ P4 from two architectures.

ability in cross-subject fall risk assessment. The combined framework of SF and DA could ensure stability across various MSST settings.

1) Sensitivity: Sensitivity is a key index in physiological classification and medical diagnosis, as it indicates the initial screening capacity for disease detection. In the case of ResNet-based results, as shown in Fig. 2 (a), although all the baselines exhibit poor sensitivities when P2 is target domain, the box plot indicates the superiority of SFDA for its high median and upper quartile. Moreover, SFDA and MhNet have relatively few sensitivities outliers (represent by dots outside the box), revealing their stability. As shown in Fig. 2 (b), sensitivity based on VggNet-16 perform similar to those based on ResNet-50 for all baselines. However, SFDA outperform other methods (except for MhNet) in terms of median, upper quartile, and the number of outliers, indicating its practicality in medical diagnosis compared to other baselines. Thus, voting a segment from multi-layer architectures may be a more sensitive way to identity high risk samples when compared to shallow pre-trained model, such as VggNet-16. To concretely record the sensitivity from each baseline, Table format of relevant results are provided in Appendix.

2) Ablation Study on SF: We evaluate the efficiency and optimal weight of SF by testing in four MSST settings, as shown in Fig. 2 (c). The different-colored polylines represent the accuracy of the proposed method under different MSST settings with varying weights. As equation (11) demonstrated, α was set from 0, 0.2, 0.4, 0.6, 0.8, 1 to explore the optimal weight. We set the loss weight of SF as 0 to represent the framework without SF. As the weight of SF increases, the overall accuracy of SFDA is getting higher based on both ResNet-50 and VggNet-16, indicating the improvement

according to subject fusion. The effectiveness of subject fusion is evident as all the model with SF perform better than that without SF. As for the optimal weight between SF and DA, SFDA generally performs better when weight of SF is larger than 0.6.

3) Distribution Discrepancy: As proxy$-\mathcal{A}-$ distance can be applied to measure the distribution distance between domains as $\text{dist}_{\mathcal{A}} = 2\,(1-2\varepsilon)$, where ε is generalization error [55], [61], we evaluate the distance between source and target domain in all MSST settings when implementing all the domain adaptation related methods. In Fig. 3 (a), compared to baselines, SFDA performs smaller $\text{dist}_{\mathcal{A}}$ for both ResNet-50 and VggNet-16, showing efficiency of reducing domains gap by SFDA. Furthermore, the overall $\text{dist}_{\mathcal{A}}$ based on VggNet-16 is smaller than that of ResNet-50 due to the feature transferability of shallow network [63]. When P2 is target domain, PAD of CDAN trained on VggNet-16 is the smallest according to Fig. 2 (b). However, it is worth noticing that the accuracy of CDAN in Table III shows that the low PAD value may result from the poor performance in identifying risk samples rather than excellent discrepancy-reducing ability. While in other MSST settings, SFDA consistently achieves the smallest PAD according to Fig. 3 (c) and (d), demonstrating its effective alleviation of domain discrepancy between source domain and target domains through subject fusion and domain adaptation.

4) Source Subject Fusion Visualization: We visualize the feature representations from the last layer in shared feature extractor through t-SNE [64]. Among them, two ResNet-based case studies with P1 and P2 as target domains (Fig. 4. (a) and (b)) and two VggNet-based case studies with P3 and P4 as target domains (Fig. 4. (c) and (d)) are presented.
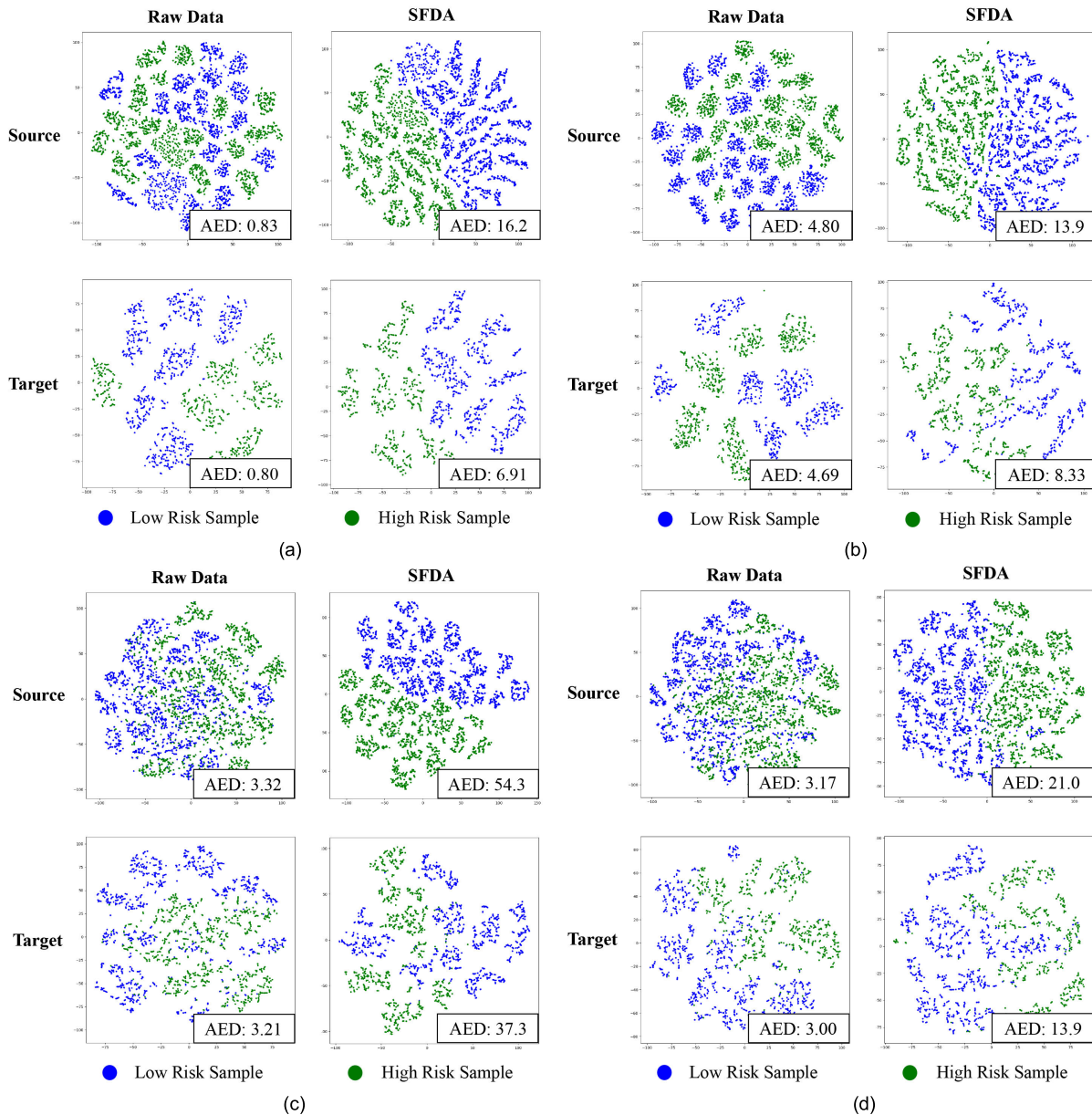
Fig. 4. t-SNE visualization of raw data and SFDA in MSST setting: (a) P2, P3, P4 → P1 in ResNet-based backbone; (b) P1, P3, P4 → P2 in ResNet-based backbone; (c) P1, P2, P4 → P3 in VggNet-based backbone; (d) P1, P2, P3 → P4 in VggNet-based backbone. The first row indicates the clustering from source domain and the second row indicates the clustering from target domain. Different colors denote different classes: high risk and low risk. Best view in color.

In general, SFDA successfully distinguishes high and low risk samples in all MSST settings, regardless of ResNet-50 or VggNet-16. Additionally, SFDA efficiently aligns the distribution regions between source and target domain, showcasing its superior generalization ability when facing novel samples. Specifically, in VggNet-based clustering shown in Fig.4 (c) and (d), the raw data distributions of low risk and high risk samples overlap severely without SFDA, resulting in poor task-specific feature extraction, which is matched with the low accuracy illustrated in Table III. Furthermore, as displayed in Fig. 4 (a) and (b), SFDA facilitates the fusion of source subject distribution, thereby alleviating the shift within source subject and enabling target samples to fully exploit source task-specific features.

To quantitatively evaluate the classification performance in feature clustering, the Average Euclidean Distance (AED) is computed from feature matrix. Within source or target domain, Euclidean Distance from each pair of high risk and low risk samples is calculated (stored as a distance matrix where each element represents the distance between a pair of high risk and low risk samples). Subsequently, dividing the sum of the distance matrices by the total number of samples yields the AED, where a higher AED indicates easier distinction between high risk and low risk samples in the source or target domain. Based on AED presented in Fig. 4, employing SFDA results in obvious higher AED, showing its efficient task-specific feature extraction and classification ability. Thus, a high AED can be a proof of high accuracy as shown in Table II and Table III.

TABLE IV
ACCURACY (%) OF DIFFERENT METHODS ON FOUR MSST SETTING BASED ON INCEPTION-V3

Source	Target	Inception	DAN	DeepCoral	DANN	CDAN	DSAN	MhNet	SFDA
P2, P3, P4	P1	73.18±0.5	75.83±1.4	72.86±1.6	75.58±1.7	74.07±2.8	77.23±2.2	**86.87±5.6**	<u>78.46±1.1</u>
P1, P3, P4	P2	51.29±0.5	52.38±1.3	51.87±1.1	51.84±0.8	54.35±1.7	54.40±0.7	**91.20±6.9**	<u>54.70±0.8</u>
P1, P2, P4	P3	69.18±0.9	72.77±1.9	71.61±0.9	70.78±1.3	<u>75.75±1.6</u>	75.42±0.9	46.00±0.7	**76.10±0.9**
P1, P2, P3	P4	72.55±1.4	72.52±0.8	73.08±1.5	76.03±2.7	**77.35±1.5**	76.32±1.6	66.95±2.4	<u>76.42±1.2</u>
Average		66.55	68.37	67.36	68.15	70.38	70.84	**72.76**	<u>71.42</u>

*The best performance is highlighted in bold, and the second-best performance is underlined.

TABLE V
ACCURACY (%) OF DIFFERENT METHODS ON FOUR MSST SETTING BASED ON RESNET-34

Source	Target	ResNet-34	DAN	DeepCoral	DANN	CDAN	DSAN	MhNet	SFDA
P2, P3, P4	P1	81.97±5.3	81.54±1.5	84.35±2.4	83.75±3.7	82.47±9.1	86.04±6.0	<u>86.87±5.6</u>	**87.92±2.6**
P1, P3, P4	P2	58.99±2.8	58.89±2.5	56.09±0.8	58.99±1.0	56.77±0.0	<u>63.29±0.2</u>	**91.20±6.9**	60.41±1.1
P1, P2, P4	P3	74.01±1.7	76.27±0.1	77.15±1.6	74.61±0.7	76.60±0.0	<u>77.33±0.3</u>	46.00±0.7	**78.55±1.1**
P1, P2, P3	P4	67.15±2.5	66.35±4.0	66.44±0.4	70.68±2.6	70.56±1.1	<u>72.04±6.5</u>	66.95±2.4	**73.74±0.4**
Average		70.53	71.13	71.01	72.01	71.60	<u>74.68</u>	72.76	**75.15**

*The best performance is highlighted in bold, and the second-best performance is underlined.

C. Results Summary and Limitations

In summary, the superior performance of SFDA, as shown in Table II and Table III, can also be explained by the t-SNE clustering and PAD results as follows: (i) The classification accuracy relies on task-specific features, which form the fundamental of proposed model. As depicted in Fig. 4, in certain MMST settings, the high risk subjects and low risk subjects could be easily distinguished in feature space, which guarantees a lower bound for fall risk assessment accuracy. (ii) Since task-specific features capturing are inevitable in MSST cross-subject setting, source subject fusion module helps reducing internal gap within source domain and prompts target samples to fully adapt to the task-specific features. (iii) In addition to the better performance in reducing domain gap (good at extracting domain-invariant features) demonstrated in Table II, Table III, Fig. 3 and Fig. 4, the lower PAD may also result from poor classification ability (bad at extracting task-specific features). For example, CDAN achieved the lowest PAD when P2 is target domain and VggNet-16 is backbone, but its accuracy in this setting is poor.

Despite the improvement achieved by SFDA, there are some limitations. Although SFDA outperforms all the baselines in average accuracy, there are a few MSST settings where SFDA fails to achieve the best performance. The outliers and larger standard deviation of sensitivity in Fig. 2 (a) and (b) of SFDA may be attributed to the challenge in fusing source subjects for two reasons. First, subjects shifts within source domain are too difficult to eliminate. Second, the coverage or representative of source subjects is limited, thereby resulting in slight fluctuation. This deficiency reveals the need for further improvement, particularly dealing with the subjects that have huge domain gaps. Moreover, the key of increasing lower bound of generalization error is that the source subjects should be representative and comprehensive enough, allowing SF to extract more task-specific features.

Based on the summary and limitations, we plan to explore the outlier identification procedure in data pre-processing to boost the overall performance. Additionally, in the future, we will investigate the combination of universal domain adaptation with subject fusion, as there are always patients with specific characteristic who are not involved in the source domain dataset. On top of that, as gait features closely connect to various diseases, it is essential to construct a new dataset containing plantar pressure data from patients with different diseases to establish the mapping between complex plantar pressure features and diseases features. In physiological data classification, other types of data, such as ECG, EEG, PPG, EMG, accelerometer data, could be used and evaluated in SFDA and related modified methods. We expect to propose a general method that can be applied in various physiological data based on the main idea of subject fusion and domain adaptation.

V. CONCLUSION

In this paper, we propose a novel method named SFDA for multi-source and single-target fall risk assessment. The proposed approach improves the generalization ability and efficiency of domain adaptation through alleviating source subject shifts. Subject fusion is achieved by adopted the adversarial module to mix the source subject distributions. By synchronously carrying out DA and SF, the robustness of adversarial network is enhanced. Our proposed approach is revealed to outperform the state-of-the-art methods in continuous plantar pressure dataset, providing a method that could be adopted in real-world clinical application.

VI. CONFLICT OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

TABLE VI
ACCURACY (%) OF DIFFERENT METHODS ON FOUR MSST SETTING BASED ON VGGNET-19

Source	Target	VggNet-19	DAN	DeepCoral	DANN	CDAN	DSAN	MhNet	SFDA
P2, P3, P4	P1	81.03±3.0	78.01±0.2	67.79±2.4	77.43±2.4	83.42±1.1	80.06±0.2	**86.87±5.6**	<u>83.72±1.1</u>
P1, P3, P4	P2	52.75±4.5	63.84±0.2	54.47±0.0	58.07±5.5	59.45±0.0	59.13±0.6	**91.20±6.9**	<u>64.69±4.2</u>
P1, P2, P4	P3	70.01±0.2	74.44±0.9	68.83±0.7	70.29±0.8	<u>74.92±0.5</u>	**76.81±0.0**	46.00±0.7	73.37±0.8
P1, P2, P3	P4	78.57±0.9	78.60±0.0	**82.27±2.4**	77.89±0.8	78.93±1.1	<u>79.69±0.0</u>	66.95±2.4	78.42±2.5
Average		70.59	73.72	68.34	70. 92	<u>74.18</u>	73.92	72.76	**75.05**

*The best performance is highlighted in bold, and the second-best performance is underlined.

TABLE VII
SENSITIVITY (%) OF DIFFERENT METHODS ON FOUR MSST SETTING BASED ON RESNET-50

Source	Target	ResNet-50	DAN	DeepCoral	DANN	CDAN	DSAN	MhNet	SFDA
P2, P3, P4	P1	77.05±11.9	80.05±1.8	89.22±4.9	88.05±2.0	87.00±0.0	82.52±0.6	**100.0±0.0**	<u>94.53±2.9</u>
P1, P3, P4	P2	54.61±2.5	53.28±0.0	56.70±0.2	<u>60.14±5.0</u>	59.34±0.0	59.87±1.9	**78.67±0.0**	59.17±3.5
P1, P2, P4	P3	81.43±4.9	83.60±0.3	79.74±0.0	83.66±3.0	85.21±0.3	82.72±1.8	<u>85.40±0.5</u>	**87.65±3.3**
P1, P2, P3	P4	77.95±0.8	**86.76±2.1**	73.10±1.1	75.62±6.4	79.40±0.0	77.87±1.7	54.99±5.0	<u>80.90±3.7</u>
Average		72.76	75.92	74.69	76.86	77.74	75.75	<u>79.76</u>	**80.56**

*The best performance is highlighted in bold, and the second-best performance is underlined.

TABLE VIII
SENSITIVITY (%) OF DIFFERENT METHODS ON FOUR MSST SETTING BASED ON VGGNET-16

Source	Target	VggNet-16	DAN	DeepCoral	DANN	CDAN	DSAN	MhNet	SFDA
P2, P3, P4	P1	64.98±12.4	69.10±8.5	56.73±0.0	53.15±4.2	66.80±0.0	73.58±4.6	**100.0±0.0**	<u>75.12±1.6</u>
P1, P3, P4	P2	44.72±7.8	54.75±1.3	47.73±6.8	32.23±5.1	57.34±0.0	<u>60.89±5.7</u>	**78.67±0.0**	53.31±4.8
P1, P2, P4	P3	73.06±3.6	78.88±0.7	<u>85.53±0.0</u>	81.39±6.6	78.80±0.4	83.10±0.6	85.40±0.5	**87.59±2.0**
P1, P2, P3	P4	**91.18±3.2**	69.11±0.0	<u>84.80±4.5</u>	83.35±2.0	78.80±0.0	68.76±0.0	54.99±5.0	84.66±4.5
Average		68.48	67.96	68.70	62.53	70.43	71.58	**79.76**	<u>75.17</u>

*The best performance is highlighted in bold, and the second-best performance is underlined.

TABLE IX
DOMAIN DIVISION IN PD DATASET

Division		Subjects
P1	PD	#31, #32, #33, #34, #35, #36, #37, #38, #39
	CO	#1, #2, #3, #4, #5, #6, #7
P2	PD	#40, #41, #42, #43, #44, #45, #46, #47, #48
	CO	#8, #9, #10, #11, #12, #13, #14
P3	PD	#49, #50, #51, #52, #53, #54, #55, #56, #57
	CO	#15, #16, #17, #18, #19, #20, #21, #22
P4	PD	#30, #58, #59, #60, #61, #62, #63, #64
	CO	#23, #24, #25, #26, #27, #28, #29

*P denotes the domain and # denotes the subject ID. As the subject ID is discontinuous, we rearranged them in order from CO to PD.

APPENDICES

In this section, we provide external validation results as well as supplementary experimental results.

A. Performance Evaluation Based on More Backbones

We additionally evaluate the robustness of SFDA based on ResNet-34, VggNet-19 and Inception-V3. As shown in Table IV, V, VI, three average classification accuracy from all MSST settings shows that SFDA outperforms other baselines regardless of the backbone architecture (Except for Inception-V3, which is only 1.34% lower than the highest accuracy achieved). In Inception-based testing, SFDA generally achieves second-best accuracy. In backbones of ResNet-34 and VggNet-19, SFDA performs the best in average accuracy. However, SFDA performed slightly different compared with ResNet-50 and VggNet-16 mainly due to the depth of network. When training with a deeper architecture, SFDA based on VggNet-19 performs almost better in all MSST settings (especially in P1, the accuracy of SFDA based on VggNet-19 has increased by 6% compared to VggNet-16). When training with a shallower architecture, SFDA based on ResNet-34 performs poorly in all MSST settings. Compared to ResNet-50, the accuracy of SFDA based on ResNet-34 in all divisions (except P3) has significantly decreased, with a maximum drop of 7.9%. However, we still could testify the robustness of SFDA upon above results and it is natural to obtain different results from different backbones due to various factors, such as network frameworks, parameters number, etc.

B. Concrete Value of Sensitivity From Fig. 2

We supplement the concrete sensitivity value from Fig.2. in Table VII and Table VIII for specifically understanding.

TABLE X
ACCURACY (%) AND SENSITIVITY (%) OF SFDA ON FOUR MSST SETTING BASED ON TWO BACKBONES

Backbone		Accuracy (%)				Sensitivity (%)			
		P1	P2	P3	P4	P1	P2	P3	P4
ResNet-34	Baseline	72.3±5.1	73.3±3.1	46.2±1.0	68.5±5.1	85.8±5.6	57.5±4.0	60.2±6.3	73.0±8.9
	SFDA	79.3±1.9	76.6±3.1	46.3±2.2	75.6±2.4	86.0±5.3	59.0±5.3	63.7±4.1	75.1±5.7
VggNet-16	Baseline	65.2±4.6	66.9±1.2	42.8±1.9	69.4±2.5	76.2±9.0	46.2±5.8	64.7±6.8	66.7±2.7
	SFDA	71.0±2.3	68.0±1.7	44.5±1.4	75.0±2.5	78.0±8.6	51.1±4.5	64.9±5.0	77.9±5.6

C. External Validation

To evaluate whether the good performance of SFDA is only specific to dataset [67] or not, we added external validation experiment on Parkinson Disease (PD) Classification based on a public PD dataset (Gait in Parkinson's Disease [52]), which is available on: https://www.physionet.org/content/gaitpdb/1.0.0. This dataset collected multichannel recordings from force sensors beneath the feet of 93 patients with Parkinson's Disease, and 73 healthy controls (CO), and detail could be checked in above website. To better implement proposed SFDA in this dataset, we split the raw data into 16×200, and each sample contained one gait cycle from 16 sensors. As the fall risk assessment dataset was built through flat walking, we therefore selected a treadmill walking study from PD dataset. Among this study, 29 CO and 35 PD patients were collected and we divide them into 4 MSST settings (4-fold cross validation) as we did before. The MSST division are summarized in Table IX. Due to the GPU memory limitation, we adopt ResNet-34 and VggNet-16 as backbones. Also, the dimensions of fully connected layer are set as 128 for less computation complexity. The accuracy of SFDA based on two backbones are shown in Table X. SFDA performed better in all MSST settings in Parkinson disease classification, which could also demonstrate that SFDA is not only specific to the fall risk assessment dataset.

D. Explanation of Equation (6)

In sensor-based human data classification, the changes in sensors' state during different experiment sessions and days [46] might significantly impact the distribution of raw data, leading to a deterioration in model prediction stability.

Hysteresis, response time, and cyclic stability are factors to evaluate the sensors' performance upon dynamic loadings. Among them, the inconsistent sensing performance between loading and unloading is caused by hysteresis behavior. Cyclic stability is considered to ensure endurance against periodic stretching and releasing cycles. However, a significant challenge for piezoresistive materials lies in their poor stability and the presence of hysteresis effects when subjected to cyclic strain loading, which causes the collected data to change over time [73]. Also, the strength of interfacial interaction severely affects the performance of the flexible sensor in long-term measurements [74], [75]. What's worse, the conductive material easily aggregates and slides, or even falls off from the matrix when subjected to long-term cyclic compressive strain, resulting in unstable sensor signal [76], [77]. Hence, the characters of sensors determine that the gap between different subjects who were collected at different days would be larger than that were collected at the same day.

In particular, the piezoresistive flexible pressure sensors [8] from the public dataset used in this paper [67] exhibit similar characters. In [8], it is observed that the conductivity of the sensor increases slowly with the number of compression times, which is caused by the decreasing resistance because of the increasing number of conductive paths in the conductive rubber during compression. After further increasing the times of compressions to six thousand (after a device has been used for a period of time), the resistance output of the sensor drops significantly, which is determined by the characteristics of the composite sensor.

To sum up, when comparing $\mathrm{Dist}(\mathbf{X}^2, \mathbf{X}^{\tilde{d}})$ (distance between source data and target data that are collected at between long time interval) and $\mathrm{Dist}(\mathbf{X}^1, \mathbf{X}^2)$ (distance between source data that are collected at same batch), it could be formulated as equation (6) in manuscript most of the time.

REFERENCES

[1] World Health Organization. (2018). *World Health Statistics 2018*. Accessed: Mar. 23, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/falls

[2] A. Ozcan, H. Donat, N. Gelecek, M. Ozdirenc, and D. Karadibak, "The relationship between risk factors for falling and the quality of life in older adults," *BMC Public Health*, vol. 5, no. 1, pp. 1–6, Dec. 2005, doi: 10.1186/1471-2458-5-90.

[3] A. F. Ambrose, G. Paul, and J. M. Hausdorff, "Risk factors for falls among older adults: A review of the literature," *Maturitas*, vol. 75, no. 1, pp. 51–61, May 2013, doi: 10.1016/j.maturitas.2013.02.009.

[4] E. E. Stone and M. Skubic, "Fall detection in homes of older adults using the Microsoft Kinect," *IEEE J. Biomed. Health Informat.*, vol. 19, no. 1, pp. 290–301, Jan. 2015, doi: 10.1109/JBHI.2014.2312180.

[5] M. Mubashir, L. Shao, and L. Seed, "A survey on fall detection: Principles and approaches," *Neurocomputing*, vol. 100, pp. 144–152, Jan. 2013, doi: 10.1016/j.neucom.2011.09.037.

[6] G. Zhao, L. Chen, and H. Ning, "Sensor-based fall risk assessment: A survey," *Healthcare*, vol. 9, no. 11, p. 1448, Oct. 2021, doi: 10.3390/healthcare9111448.

[7] S. J. Preece, J. Y. Goulermas, L. P. J. Kenney, and D. Howard, "A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 3, pp. 871–879, Mar. 2009, doi: 10.1109/tbme.2008.2006190.

[8] D. Wang, J. Ouyang, P. Zhou, J. Yan, L. Shu, and X. Xu, "A novel low-cost wireless footwear system for monitoring diabetic foot patients," *IEEE Trans. Biomed. Circuits Syst.*, vol. 15, no. 1, pp. 43–54, Feb. 2021, doi: 10.1109/TBCAS.2020.3043538.

[9] S. Liang, Y. Liu, G. Li, and G. Zhao, "Elderly fall risk prediction with plantar center of force using ConvLSTM algorithm," in *Proc. IEEE Int. Conf. Cyborg Bionic Syst. (CBS)*, Sep. 2019, pp. 36–41, doi: 10.1109/CBS46900.2019.9114487.

[10] S. Wu et al., "MhNet: Multi-scale spatio-temporal hierarchical network for real-time wearable fall risk assessment of the elderly," *Comput. Biol. Med.*, vol. 144, May 2022, Art. no. 105355, doi: 10.1016/j.compbiomed.2022.105355.

[11] C. Tunca, G. Salur, and C. Ersoy, "Deep learning for fall risk assessment with inertial sensors: Utilizing domain knowledge in spatio-temporal gait parameters," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 7, pp. 1994–2005, Jul. 2020, doi: 10.1109/JBHI.2019.2958879.

[12] B. M. Meyer et al., "Wearables and deep learning classify fall risk from gait in multiple sclerosis," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 5, pp. 1824–1831, May 2021, doi: 10.1109/JBHI.2020.3025049.

[13] M. Martinez and P. L. De Leon, "Falls risk classification of older adults using deep neural networks and transfer learning," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 1, pp. 144–150, Jan. 2020, doi: 10.1109/JBHI.2019.2906499.

[14] J. Howcroft, J. Kofman, and E. D. Lemaire, "Review of fall risk assessment in geriatric populations using inertial sensors," *J. NeuroEng. Rehabil.*, vol. 10, no. 1, pp. 1–12, 2013, doi: 10.1186/1743-0003-10-91.

[15] X. Fan et al., "Automatic fall risk assessment with Siamese network for stroke survivors using inertial sensor-based signals," *Int. J. Intell. Syst.*, vol. 37, no. 9, pp. 6168–6184, Sep. 2022, doi: 10.1002/int.22838.

[16] K. McManus, B. R. Greene, L. G. M. Ader, and B. Caulfield, "Development of data-driven metrics for balance impairment and fall risk assessment in older adults," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 7, pp. 2324–2332, Jul. 2022, doi: 10.1109/TBME.2022.3142617.

[17] B. R. Greene, I. Premoli, K. Mcmanus, D. McGrath, and B. Caulfield, "Predicting fall counts using wearable sensors: A novel digital biomarker for Parkinson's disease," *Sensors*, vol. 22, no. 1, p. 54, Dec. 2021, doi: 10.3390/s22010054.

[18] A. B. A. Pinto, G. A. de Assis, L. C. B. Torres, T. Beltrame, and D. M. G. Domingues, "Wearables and detection of falls: A comparison of machine learning methods and sensors positioning," *Neural Process. Lett.*, vol. 54, no. 3, pp. 2165–2179, Jun. 2022, doi: 10.1007/s11063-021-10724-2.

[19] D. Giansanti, G. Maccioni, S. Cesinaro, F. Benvenuti, and V. Macellari, "Assessment of fall-risk by means of a neural network based on parameters assessed by a wearable device during posturography," *Med. Eng. Phys.*, vol. 30, no. 3, pp. 367–372, Apr. 2008, doi: 10.1016/j.medengphy.2007.04.006.

[20] L. Zhang and X. Gao, "Transfer adaptation learning: A decade survey," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jun. 21, 2022, doi: 10.1109/TNNLS.2022.3183326.

[21] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010, doi: 10.1109/tkde.2009.191.

[22] M. Wang and W. Deng, "Deep visual domain adaptation: A survey," *Neurocomputing*, vol. 312, pp. 135–153, Oct. 2018, doi: 10.1016/j.neucom.2018.05.083.

[23] S. Zhao et al., "A review of single-source deep unsupervised visual domain adaptation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 2, pp. 473–493, Feb. 2022, doi: 10.1109/TNNLS.2020.3028503.

[24] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Trans. Neural Netw.*, vol. 22, no. 2, pp. 199–210, Feb. 2011, doi: 10.1109/TNN.2010.2091281.

[25] M. Long, Y. Cao, J. Wang, and M. I. Jordan, "Learning transferable features with deep adaptation networks," in *Proc. 32nd Int. Conf. Mach. Learn.*, vol. 37, 2015, pp. 97–105.

[26] M. Long et al., "Deep transfer learning with joint adaptation networks," in *Proc. ICML*, 2017, pp. 2208–2217.

[27] B. Sun and K. Saenko, "Deep coral: Correlation alignment for deep domain adaptation," in *Proc. ECCV*, 2016, pp. 443–450, doi: 10.1007/978-3-319-49409-8_35.

[28] G. Kang, L. Jiang, Y. Yang, and A. G. Hauptmann, "Contrastive adaptation network for unsupervised domain adaptation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4888–4897.

[29] J. Li, E. Chen, Z. Ding, L. Zhu, K. Lu, and H. T. Shen, "Maximum density divergence for domain adaptation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 11, pp. 3918–3930, Nov. 2021, doi: 10.1109/TPAMI.2020.2991050.

[30] R. Li et al., "Adversarial domain adaptation of asymmetric mapping with CORAL alignment for intelligent fault diagnosis," *Meas. Sci. Technol.*, vol. 33, no. 5, May 2022, Art. no. 055101, doi: 10.1088/1361-6501/ac3d47.

[31] M. Xu et al., "Adversarial domain adaptation with domain mixup," in *Proc. AAAI Conf. Artif. Intell.*, Apr. 2020, vol. 34, no. 4, pp. 6502–6509, doi: 10.1609/aaai.v34i04.6123.

[32] Y. Ganin et al., "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2030–2096, 2017.

[33] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2962–2971.

[34] A. Kumar et al., "Co-regularized alignment for unsupervised domain adaptation," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 9367–9378.

[35] C. Yu, J. Wang, Y. Chen, and M. Huang, "Transfer learning with dynamic adversarial adaptation network," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2019, pp. 778–786, doi: 10.1109/ICDM.2019.00088.

[36] X. Zhang, L. Wu, X. Zhang, C. Chen, C. Li, and X. Chen, "Multi-source domain generalization and adaptation toward cross-subject myoelectric pattern recognition," *J. Neural Eng.*, vol. 20, no. 1, Feb. 2023, Art. no. 016050, doi: 10.1088/1741-2552/acb7a0.

[37] S. Zhao, B. Li, C. Reed, P. Xu, and K. Keutzer, "Multi-source domain adaptation in the deep learning era: A systematic survey," 2020, arXiv:2002.12169.

[38] R. Matthew et al., "Learning to learn without forgetting by maximizing transfer and minimizing interference," in *Proc. ICLR*, 2019, pp. 1–31.

[39] Y. Dai et al., "Adversarial training based multi-source unsupervised domain adaptation for sentiment analysis," in *Proc. AAAI*, 2020, pp. 7618–7625, doi: 10.1609/aaai.v34i05.6262.

[40] J. Guo, D. Shah, and R. Barzilay, "Multi-source domain adaptation with mixture of experts," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2018, pp. 1–11.

[41] F. Deng, S. Tu, and L. Xu, "Multi-source unsupervised domain adaptation for ECG classification," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2021, pp. 854–859, doi: 10.1109/BIBM52615.2021.9669755.

[42] X. Yang, C. Deng, T. Liu, and D. Tao, "Heterogeneous graph attention network for unsupervised multiple-target domain adaptation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 4, pp. 1992–2003, Apr. 2022, doi: 10.1109/TPAMI.2020.3026079.

[43] M. Karimpour, S. N. Saray, J. Tahmoresnezhad, and M. P. Aghababa, "Multi-source domain adaptation for image classification," *Mach. Vis. Appl.*, vol. 31, no. 6, pp. 1–19, Sep. 2020, doi: 10.1007/s00138-020-01093-2.

[44] P. Di et al., "Fall detection and prevention control using walking-aid cane robot," *IEEE/ASME Trans. Mechatronics*, vol. 21, no. 2, pp. 625–637, Apr. 2016, doi: 10.1109/TMECH.2015.2477996.

[45] Z. Song et al., "Fall risk assessment for the elderly based on weak foot features of wearable plantar pressure," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 1060–1070, 2022, doi: 10.1109/TNSRE.2022.3167473.

[46] B.-Q. Ma, H. Li, Y. Luo, and B.-L. Lu, "Depersonalized cross-subject vigilance estimation with adversarial domain generalization," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2019, pp. 1–8, doi: 10.1109/IJCNN.2019.8852347.

[47] X. Zhang, X. Zhang, L. Wu, C. Li, X. Chen, and X. Chen, "Domain adaptation with self-guided adaptive sampling strategy: Feature alignment for cross-user myoelectric pattern recognition," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 1374–1383, 2022, doi: 10.1109/TNSRE.2022.3173946.

[48] J. Hoffman et al., "Cycada: Cycle-consistent adversarial domain adaptation," in *Proc. ICML*, vol. 80, 2018, pp. 1989–1998.

[49] S. Rakshit et al., "Unsupervised multi-source domain adaptation driven by deep adversarial ensemble learning," in *Proc. GCPR*, 2019, pp. 485–498.

[50] Y. Zhu et al., "Deep subdomain adaptation network for image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 4, pp. 1713–1722, Apr. 2021, doi: 10.1109/TNNLS.2020.2988928.

[51] J. Yosinski et al., "How transferable are features in deep neural networks?" in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 1–9.

[52] A. L. Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. 215–220, Jun. 2000.

[53] K. M. Borgwardt, A. Gretton, M. J. Rasch, H.-P. Kriegel, B. Schölkopf, and A. J. Smola, "Integrating structured biological data by kernel maximum mean discrepancy," *Bioinformatics*, vol. 22, no. 14, pp. 49–57, Jul. 2006, doi: 10.1093/bioinformatics/btl242.

[54] S. Ben-David, J. Blitzer, K. Crammer, and F. Pereira, "Analysis of representations for domain adaptation," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 19. 2007, pp. 137–144.

[55] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, "A theory of learning from different domains," *Mach. Learn.*, vol. 79, nos. 1–2, pp. 151–175, May 2010, doi: 10.1007/s10994-009-5152-4.

[56] K. Crammer, M. Kearns, and J. Wortman, "Learning from multiple sources," *J. Mach. Learn. Res.*, vol. 9, no. 8, pp. 1–18, 2008, doi: 10.1109/AAMAS.2004.159.

[57] R. Xu, Z. Chen, W. Zuo, J. Yan, and L. Lin, "Deep cocktail network: Multi-source unsupervised domain adaptation with category shift," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3964–3973, doi: 10.1109/CVPR.2018.00417.

[58] Y. Li et al., "Extracting relationships by multi-domain matching," in *Proc. NIPS*, vol. 31, 2018, pp. 1–12.

[59] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[60] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[61] Y. Mansour, M. Mohri, and A. Rostamizadeh, "Domain adaptation: Learning bounds and algorithms," in *Proc. COLT*, 2009, pp. 1–16.

[62] M. Long et al., "Conditional adversarial domain adaptation," in *Proc. NIPS*, vol. 31, 2018, pp. 1–11, doi: 10.1007/978-3-030-01237-3_9.

[63] M. Long, Y. Cao, Z. Cao, J. Wang, and M. I. Jordan, "Transferable representation learning with deep adaptation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 12, pp. 3071–3085, Dec. 2019, doi: 10.1109/TPAMI.2018.2868685.

[64] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

[65] J. Wang, Y. Chen, S. Hao, W. Feng, and Z. Shen, "Balanced distribution adaptation for transfer learning," in *Proc. ICDM*, 2017, pp. 1129–1134, doi: 10.1109/ICDM.2017.150.

[66] L. Shu et al., "Monitoring diabetic patients by novel intelligent footwear system," in *Proc. Int. Conf. Computerized Healthcare (ICCH)*, Dec. 2012, pp. 91–94, doi: 10.1109/ICCH.2012.6724478.

[67] G. Hu et al., "A dataset for falling risk assessment of the elderly using wearable plantar pressure," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2022, pp. 2900–2904, doi: 10.1109/BIBM55620.2022.9995052.

[68] H. Wang, W. Yang, Z. Lin, and Y. Yu, "TMDA: Task-specific multi-source domain adaptation via clustering embedded adversarial training," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2019, pp. 1372–1377, doi: 10.1109/ICDM.2019.00176.

[69] H. Zhao et al., "Adversarial multiple source domain adaptation," in *Proc. NIPS*, vol. 31, 2018, pp. 1–12.

[70] J. Fan et al., "Unsupervised domain adaptation by statistics alignment for deep sleep staging networks," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 205–216, 2022, doi: 10.1109/TNSRE.2022.3144169.

[71] D. Liu, J. Zhang, H. Wu, S. Liu, and J. Long, "Multi-source transfer learning for EEG classification based on domain adversarial neural network," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 218–228, 2023, doi: 10.1109/TNSRE.2022.3219418.

[72] F. Wei, X. Xu, T. Jia, D. Zhang, and X. Wu, "A multi-source transfer joint matching method for inter-subject motor imagery decoding," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 1258–1267, 2023, doi: 10.1109/TNSRE.2023.3243257.

[73] T. Yang, D. Xie, Z. Li, and H. Zhu, "Recent advances in wearable tactile sensors: Materials, sensing mechanisms, and device performance," *Mater. Sci. Eng., R, Rep.*, vol. 115, pp. 1–37, May 2017, doi: 10.1016/j.mser.2017.02.001.

[74] M. Zhu et al., "A mechanically interlocking strategy based on conductive microbridges for stretchable electronics," *Adv. Mater.*, vol. 34, no. 7, Feb. 2022, doi: 10.1002/adma.202101339.

[75] S. Pan et al., "Mechanically interlocked hydrogel-elastomer hybrids for on-skin electronics," *Adv. Funct. Mater.*, vol. 30, no. 29, Jul. 2020, Art. no. 1909540, doi: 10.1002/adfm.201909540.

[76] Z. Sang, K. Ke, and I. Manas-Zloczower, "Design strategy for porous composites aimed at pressure sensor application," *Small*, vol. 15, no. 45, Nov. 2019, Art. no. 1903487, doi: 10.1002/smll.201903487.

[77] Y. Li, P. Cui, L. Wang, H. Lee, K. Lee, and H. Lee, "Highly bendable, conductive, and transparent film by an enhanced adhesion of silver nanowires," *ACS Appl. Mater. Interfaces*, vol. 5, no. 18, pp. 9155–9160, Sep. 2013, doi: 10.1021/am402578d.

[78] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826, doi: 10.1109/CVPR.2016.308.

[79] D. Jung et al., "Deep neural network-based gait classification using wearable inertial sensor data," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 3624–3628.

[80] R. Wang, H. Jiang, Z. Li, and Y. Liu, "A deep ensemble learning model for rolling bearing fault diagnosis," in *Proc. IEEE Int. Conf. Prognostics Health Manag. (ICPHM)*, Jun. 2022, pp. 133–136.