MSSDA: Multi-Sub-Source Domain Adaptation for Diabetic Foot Neuropathy Recognition

Yan Zhong[1], Zhixin Yan[1], Yi Xie[1], Shibin Wu[2], Huaidong Zhang[1*], Lin Shu [1*], Peiru Zhou[3]

[1]South China University of Technology
[2]City University of Hong Kong
[3]The Fifth Affiliated Hospital of Jinan University

Abstract

Diabetic foot neuropathy (DFN) is a critical factor leading to diabetic foot ulcers, which is one of the most common and severe complications of diabetes mellitus (DM) and is associated with high risks of amputation and mortality. Despite its significance, existing datasets do not directly derive from plantar data and lack continuous, long-term foot-specific information. To advance DFN research, we have collected a novel dataset comprising continuous plantar pressure data to recognize diabetic foot neuropathy. This dataset includes data from 94 DM patients with DFN and 41 DM patients without DFN. Moreover, traditional methods divide datasets by individuals, potentially leading to significant domain discrepancies in some feature spaces due to the absence of mid-domain data. In this paper, we propose an effective domain adaptation method to address this proplem. We split the dataset based on convolutional feature statistics and select appropriate sub-source domains to enhance efficiency and avoid negative transfer. We then align the distributions of each source and target domain pair in specific feature spaces to minimize the domain gap. Comprehensive results validate the effectiveness of our method on both the newly proposed dataset for DFN recognition and an existing dataset.

Introduction

Diabetes mellitus (DM) is a serious non-communicable disease and a global public health issue. According to the International Diabetes Federation (Cho et al. 2018), approximately 537 million adults worldwide had diabetes in 2021, and this number is projected to reach 693 million by 2045. Over 30% of diabetic patients will develop diabetic peripheral neuropathy (DPN), with the incidence increasing with age (van Schie 2008; Carls et al. 2011). DPN affects the autonomic, sensory, and motor nervous systems, compromising skin integrity and sensation in the feet, thereby increasing susceptibility to injury and diabetic foot ulcers (DFU) (Fernando et al. 2013).

Early screening and proactive management can prevent 45% to 85% of foot ulcers (Association et al. 1999). Researches (Coppini et al. 2001; Kastenbauer et al. 2001) indicate that uneven plantar pressure distribution, particularly high pressure in certain areas, significantly contributes to foot ulcers, with a correlation of 70% to 90%. These studies have spurred research into wearable footwear to monitor plantar pressure data (Wang et al. 2020; Sazonov, Hegde, and Tang 2013) and analyze real-time changes to screen for peripheral neuropathy early, which is crucial for preventing foot ulcers. Studies on plantar pressure abnormality detection have achieved high accuracy in intra-subject setting using artificial neural networks (Wafai et al. 2014), SVM (Botros et al. 2016) and backpropagation neural network (Liu 2017).

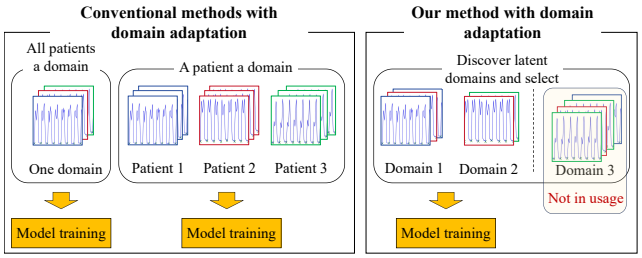

Figure 1: Unlike conventional methods, our method aims to discover latent domains and use the source data with some specific domain labels to train the model but not all of them.

DFN is a is a specific form of DPN, primarily affecting the nervous system of the feet in diabetic patients. In the field of diabetic foot neuropathy recognition, however, there is no dataset specifically focused on foot information. Some datasets are constructed from electronic health records, including age and pain-related medication prescriptions (DuBrava et al. 2017) while others include more complex indicators such as patient medical history, physical examination and biochemical test results (Lian et al. 2023). All these datasets are not directly related to foot information, lacking continuous, long-term foot-specific information. These issues motivat us to create a dataset for DFN recognition collected using wearable shoes, which contains more foot-specific information. We name it DFN-DS, which includes 5-minute continuous plantar pressure data from 94 DM patients with DFN and 41 DM patients without DFN.

Although many methods perform well under intra-subject settings, they often perform poorly under cross-subject settings due to the significant data distribution gap between in-

*Corresponding author:{huaidongz, shul}@scut.edu.cn

dividuals. Transfer learning methods are introduced to reduce this gap in biomedical signal processing. In ECG tasks, MS-MDA (Chen et al. 2021) achieve good performance in emotion recognition. In fall risk assessment tasks, MhNet is proposed based on adversarial domain adaptation, achieving good performance with additional few-step setting (Wu et al. 2022). However, these methods either divide the dataset by subject for multi-source domain adaptation or combine all samples for single-source domain adaptation. We believe the former may fail due to the significant domain gap caused by the absence of mid-domain data to bridge the source and target domain, while the latter does not leverage the advantages of multi-source domain adaptation.

In this paper, we propose a new three-stage alignment framework to overcome these issues. The first stage trains a model to separate all the samples as well as possible using contrastive learning. In the second stage, the original source dataset is divided into K sub-source domains by convolutional feature statistics, where K is determined by the Bayesian Information Criterion. Source samples are then assigned pseudo domain labels. In the third stage, we select the source samples with some proper domain labels according to a strategy to avoid negative transfer, and then we align the distributions of each pair of source and target domains in multiple feature spaces.

The contributions of this paper are summarized as follows: (1) We propose a continuous plantar pressure dataset, the first constructed from plantar pressure data for diabetic foot neuropathy recognition. (2) We propose a novel framework for biomedical signal processing that divides the dataset by convolutional feature statistics and selects some proper sub-source domains, then aligns the distributions of each pair of source and target domains in multiple feature spaces. (3) We conduct comprehensive experiments on two datasets, validating the effectiveness of the proposed model through experimental results.

Related Work

Single-source Domain Adaptation (SDA)

SDA aims to reduce the domain gap in the feature space when data may follow different distributions, which often leads to poor performance of traditional methods. Based on the generalization bound (Ben-David et al. 2006, 2010) measured by Maximum Mean Discrepancy (MMD), DAN (Long et al. 2015) is proposed to mitigate the shift in feature space. Deep CORAL (Sun and Saenko 2016) calculate the distribution gap between source domain and target domain using second-order statistics instead of MMD. Subsequently, JAN (Long et al. 2017) is developed to address the joint distribution gap. MCD (Saito et al. 2018) is then proposed to approximate the disparity difference in the bound by the disagreement between two classifiers' outputs. Domain-adversarial methods are developed with the emergence of Generative Adversarial Network (GAN). DANN (Ganin et al. 2016) is the first to propose adversarial domain adaptation based on the generalization bound and the adversarial idea. CDAN (Long et al. 2018) advances the theoretical underpinnings with Disparity Discrepancy, pushing the boundaries of domain adaptation methods. Inspired by these methods, MhNet are proposed and arrive good performance in falling risk assessment tasks (Wu et al. 2022).

Multi-source Domain Adaptation (MDA)

Unlike single-source domain adaptation, multi-source domain adaptation involves multiple source domains, introducing more complex inter-domain gaps. DSAN (Zhu et al. 2020) divides the dataset into several domains, considering data with the same category labels to share the same domain label. M3SDA (Peng et al. 2019) aligns multiple source domains with the target domain while also ensuring alignment among all the source domains, which aims to unify data from different domains within a common feature space. In contrast, MFSAN (Zhu, Zhuang, and Wang 2019) aligns the distributions of each pair of source and target domains in multiple feature spaces, processing N alignments simultaneously for N source domains and a target domain.

Many works in the field of biomedical signal process are related to MDA methods, such as gait analysis, ECG tasks, EMG tasks and EEG tasks. Usually in their methods, datasets are divided by individual, with each individual's data constituting a separate domain. Some approaches align multi-source domains and a target domain in a feature spaces using the idea of M3SDA (Wu et al. 2023; Mu et al. 2020). Some others process alignment with the idea of MFSAN, meaning that they try to align multi-source domains and a target domain in multi feature spaces (Chen et al. 2021; Deng, Tu, and Xu 2021; She et al. 2023; Guo, Gu, and Yang 2021).

However, these methods are typically designed for datasets with fewer than 50 subjects, most having less than 20. As the number of subjects increases to 100 or even 500, their approaches require substantial computational resources and time due to cross-domain computations and a parallel architecture based on the number of subjects. Moreover, using data from all individuals without selection can lead to negative transfer, while the distribution gap can be hard to reduce because of the absence of mid-domain data to bridge source domain and target domain under a patient a domain setting. To address these problems, we propose to split datasets based on convolutional feature statistics, rather than relying on subjects as in conventional methods. Additionally, our method incorporates a process for selecting appropriate subsource domains to avoid negative transfer, ensuring that not all source domains are considered for alignment with the target domain. We then leverage the idea of aligning the distributions of each pair of source and target domains across multiple feature spaces to enhance performance.

Proposed Method

Considering that we have a source domain constructed by N individuals' data $\mathcal{D}_S = \{\mathcal{D}_1, \mathcal{D}_2, ..., \mathcal{D}_N\}$, an target domain constructed by one testing individual's data $\mathcal{D}_T$ and total dataset $\mathcal{D}_{Total} = \mathcal{D}_S \cup \mathcal{D}_T$. Note that $\mathcal{D}_S \cap \mathcal{D}_T = \emptyset$. We get labeled source samples $\{(x_i^s, y_i^s)\}_{i=1}^{|\mathcal{D}_S|}$ from $\mathcal{D}_S$, where $|\mathcal{D}_S|$ refer to the total number of samples in $\mathcal{D}_S$. Similarly we have unlabelled target samples $\{x_j^t\}_{j=1}^{|\mathcal{D}_T|}$ from target domain

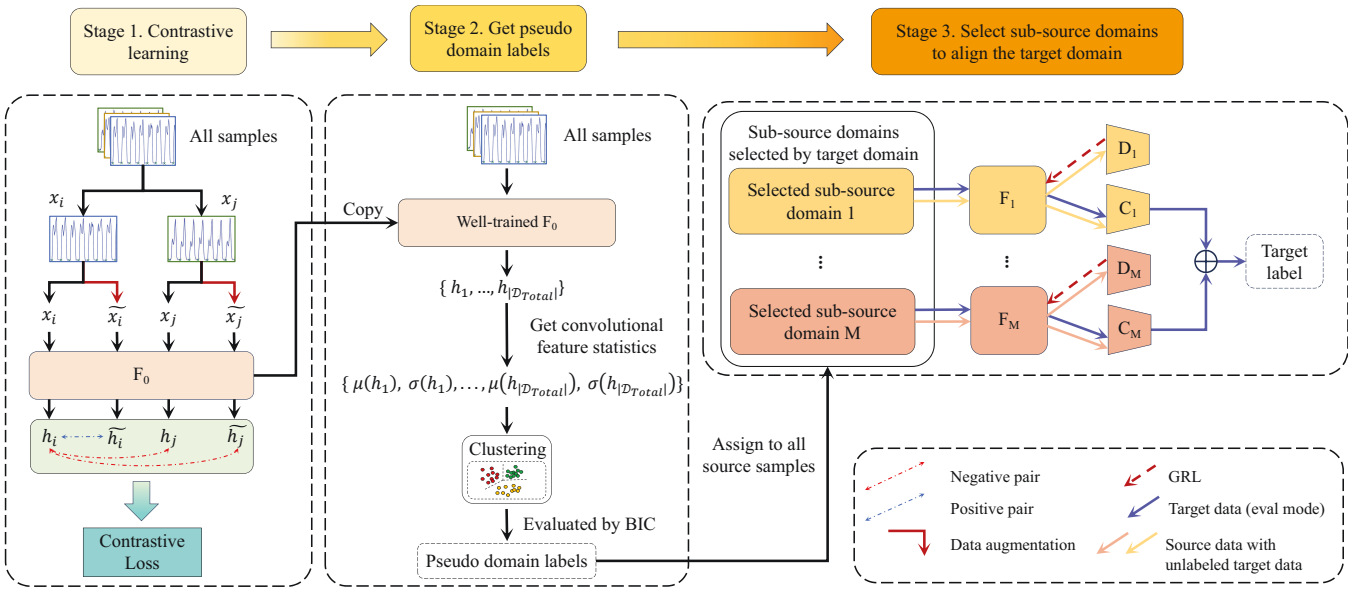

Figure 2: An overview of the proposed three-stage framework MSSDA. F_i denotes a feature extractor, D_i signifies a domain discriminator, C_i represents a classifier. Note that the F_0 trained in stage 1 is used in stage 2 while F_1,..., F_M are not fine-tuned from F_0. In stage 3, there are specific feature extractors, domain discriminator and classifiers for each selected source domain. Please be aware that F_0, F_1,..., F_M neither share the network architecture nor the weights. (Best viewed in color.)

$\mathcal{D}_T$ and all unlabeled samples $\{x_m\}_{m=1}^{|\mathcal{D}_{Total}|}$ from $\mathcal{D}_{Total}$. Our goal is to mitigate the domain shift between $\mathcal{D}_S$ and $\mathcal{D}_T$ to learn common domain-invariant features in order to improve prediction accuracy in the target domain.

Three-Stage Alignment Framework

Our framework consists of three stages as shown in Figure 2. In the first stage, we try to make all samples distinguishable from each other in the feature space by contrastive learning. Then in the second stage, we feed all samples to the well-trained feature extractor from the first stage and get the outputs, then we calculate the convolutional feature statistics (mean and standard deviation) of the output. After performing Gaussian Mixture Model (GMM) on the convolutional feature statistics, we obtain pseudo domain labels and assign corresponding ones to all source samples. The number of clusters is determined using the Bayesian Information Criterion (BIC). Finally, in the third stage, we select M sub-source domains that are closest to the target domain and align the distributions of each pair of source and target domain in multiple feature spaces.

Stage 1. Contrastive Learning Here we show the details in stage 1. Contrastive learning is widely used as a pre-training method to enhance the feature extractor's ability to learn effective representations in the field of biomedical signal process (Wang et al. 2024; Lai et al. 2023). The main idea of it is to mine data consistency by bringing similar data (or positive pair) closer together and pushing dissimilar data (or negative pair) further apart. Unlike previous methods, we do not employ contrastive learning as a pre-training technique. Instead, we harness its influence within the feature space to widen the distance between all samples whatever from $\mathcal{D}_T$ or $\mathcal{D}_S$, thereby enhancing the clustering effect in the second stage, which will be shown in detail below. Specifically, we set $(x_i, \tilde{x}_i)$ as a positive pair, (x_i, x_j) and $(x_i, \tilde{x}_j)$ as negative pairs, where $\tilde{x}_i$ is the augmented view of x_i, $i \neq j$ and $x_i, x_j \in \mathcal{D}_{Total}$. We then make all samples distinguishable from each other in the feature space using contrastive loss, defined as

$$\mathcal{L}_{\mathcal{C}} = \mathbb{E}\left[-\log \frac{\exp(h_i \cdot \tilde{h}_i)}{\sum_{j=1}^{J} \exp(h_i \cdot \tilde{h}_j) + \mathbb{1}_{[i\neq j]} \exp(h_i \cdot h_j)}\right], \tag{1}$$

where $h_i = F_0(x_i), x_i \in \mathcal{D}_{Total}$, $J = |\mathcal{D}_{Total}|$, F_0 stands for the feature extractor used in this stage, $\cdot$ means dot product and $\mathbb{1}_{[i\neq j]}$ stands for an indicator function that equals 1 when $i \neq j$ and 0 other wise. Note that the feature extractor F_0 well-trained in this stage will be used in following stages.

Stage 2. Clustering and Get Pseudo Domain Labels Inspired by Matsuura and Harada's work (Matsuura and Harada 2020), we assume that the latent domains of data are reflected in their styles, specifically in the convolutional feature statistics (mean and standard deviations). Therefore, we firstly obtain convolutional feature statistics from the well-trained feature extractor F_0. Specifically, we calculate the mean and standard deviations of the source samples by channel c as follows:

$$\mu_c(F_0(x_i)) = \frac{1}{HW} \sum_{h=1}^{H} \sum_{w=1}^{W} (F_0(x_i))_{chw}, \tag{2}$$

$$\sigma_c(F_0(x_i)) = \sqrt{\frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W}(F_0(x_i)_{chw} - \mu_c(F_0(x_i)))^2}, \tag{3}$$

where $x_i \in \mathcal{D}_{Total}$, c, h, w respectively refer to channel, height and width of the representation of x_i transformed by well-trained F_0 in the feature space.

We represent them in a simple concise form in Figure 2, in which $h_i = F_0(x_i)$, $\mu(h_i) = \{\mu_c(F_0(x_i)\}_{c=1}^{C}$, $\sigma(h_i) = \{\sigma_c(F_0(x_i)\}_{c=1}^{C}$. Then we utilize GMM as the clustering method on these convolutional feature statistics. Finally, the influence of different cluster numbers will be evaluated by the Bayesian Information Criterion (BIC) (Schwarz 1978), motivated by SSDA (Lu et al. 2021),

$$BIC = \text{-}2\ln L + k\ln m, \tag{4}$$

where L represents the maximized value of the likelihood function for the estimated model, k represents the number of free parameters to be estimated, and m is the sample size. We seek proper cluster number K which minimizes BIC.

Using BIC, we determine a certain number of clusters, which differs from Matsuura and Harada's approach. Additionally, they utilize a stack of convolutional feature statistics obtained from lower layers of the feature extractor, whereas we choose those from the last layer because it is task-specific that separate the data as effectively as possible, aligning with our objectives.

In the end of stage 2, after assigning K types of numerical pseudo-domain labels to source samples, we get K sub-source domains and K cluster centers, denoted as $\mathcal{D}_{sub_k}$, $Center_{sub_k}, k \in \{1, ..., K\}$, respectively.

Stage 3. Select Sub-Source Domains to Align the Target Domain Compared with the previous works in the field of biomedical signal process, our method do not use all the source domains, instead, we select some of them to avoid negative transfer which may be happened as the individual number of the dataset increases. We select M sub-source domains by the distance between the sub-source domain center $Center_{sub_k}$ and the target samples, which is calculated as followed:

$$dis_k = \max_{x_j^t \sim \mathcal{D}_T} \left\| T(\mu(F_0(x_j^t)), \sigma(F_0(x_j^t))), Center_{sub_k} \right\|_2, \tag{5}$$

where T stands for PCA with dim = 2 and $\|\cdot\|_2$ refers to L_2 norm. Then we select M ($M \leq K$) sub-source domains by the distance calculated in Eq. (5):

$$\mathcal{D}_{ssub_1} = \underset{k \in \{1,...,K\}}{\operatorname{argmin}} dis_k, \tag{6}$$

$$\mathcal{D}_{ssub_M} = \underset{k \in \{1,...,K\} \setminus \{\mathcal{D}_{ssub_1}, ..., \mathcal{D}_{ssub_{M-1}}\}}{\operatorname{argmin}} dis_k, \tag{7}$$

where M is determined by user. We name them selected sub-source domain 1,..., selected sub-source domain M, respectively, as shown in Figure 2. Then we try to reduce the domain gap between these M sub-source domains and target domain in order to improve cross-subjects performance. We utilize the idea that aligns the distributions of each pair of source and target domains in multiple feature spaces (Zhu, Zhuang, and Wang 2019). In a traditional domain-adversarial framework, such as DANN (Ganin et al. 2016), there are three parts, a feature extractor F to get domain-invariant features, a domain discriminator D to distinguish which domain the output of the feature extractor is from and a classifier C to predict the category label of the output of the feature extractor. By loss function, F is encouraged to extract the features that are challenging for D to distinguish, while the D is trained to correctly predict the domain label of the output of the feature extractor. This creates an adversarial relationship between F and D, allowing F to extract domain-invariant features across different domains. The framework we used in stage 3 can be seen as a parallel architecture of M DANNs. The target domain is aligned with each source domain in a specified sub-networks, as showed in Figure 2. Loss function is calculated as bellow:

$$\mathcal{L}_{\text{cls}} = \sum_{k=1}^{M} \sum_{(x_i^s, y_i^s) \sim \mathcal{D}_{ssub_k}} \mathcal{L}_{CE}(C_k(F_k(x_i^s)), y_i^s), \tag{8}$$

$$\mathcal{L}_{\text{adv}} = \sum_{k=1}^{M} \sum_{\substack{x_i^s \sim \mathcal{D}_{ssub_k}, \\ x_j^t \sim \mathcal{D}_T}} \mathcal{L}_{CE}(D_k(F_k(x_i^s)), 1) + \mathcal{L}_{CE}(D_k(F_k(x_j^t)), 0), \tag{9}$$

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{cls}} - \alpha\mathcal{L}_{\text{adv}}, \tag{10}$$

where $\mathcal{L}_{\text{CE}}$ denotes the use of cross-entropy (Xie et al. 2024, 2023), $\mathcal{L}_{\text{cls}}$ represents the category classification loss, $\mathcal{L}_{\text{adv}}$ refers to the adversarial loss, α is a trade-off parameter and $\mathcal{L}_{\text{total}}$ is the total loss. Finally, to predict the labels of target samples, we compute the average of all classifier outputs.

Experiments

We evaluate our proposed method (MSSDA) with some transfer learning methods that are widely used in the field of biomedical signal process on two datasets: one is our proposed dataset DFN-DS and the other one is Fall Risk Assessment dataset (Hu et al. 2022), both involving plantar pressure data. By using these two dataset, we verify whether our method is suitable for biomedical signal datasets. Our code and dataset DFN-DS will be available at https://github.com/YuGuilliman/MSSDA.

Data Preparation

Diabetic Foot Neuropathy Dataset (DFN-DS) It includes plantar pressure data from 135 subjects, with 94 diabetic patients having DFN (labeled as 1) and 41 without DFN (labeled as 0). We recorded data as each patient walked freely in a straight line for 5 minutes. After data cleaning we obtained 6,983 samples. Data has the shape $x \in \mathbb{R}^{147\times16}$,

Algorithm 1: Training algorithm

Require: source samples $\{(x_i^s, y_i^s)\}_{i=1}^{|\mathcal{D}_S|}$, target data $\{x_j^t\}_{j=1}^{|\mathcal{D}_T|}$, trade-off parameter α, and batch size bs

Initialize parameters of F_0, F_1, F_2, D_1, D_2, C_1, C_2
Train F_0 by Eq.(1) using all samples $\{x_m\}_{m=1}^{|\mathcal{D}_{Total}|} = \{x_i^s\}_{i=1}^{|\mathcal{D}_S|} \cup \{x_j^t\}_{j=1}^{|\mathcal{D}_T|}$
Collect the representation $\{F_0(x_m)\}_{m=1}^{|\mathcal{D}_{Total}|}$ of all the samples and get their convolutional feature statistics $\{\mu(h_m), \sigma(h_m)\}_{m=1}^{|\mathcal{D}_{Total}|}$ by Eq.(2) (3), then use GMM as cluster method on these convolutional feature statistics and get the proper cluster number K evaluated by BIC
Assign pseudo domain labels of K types to all source samples
Select M ($M \leq K$) proper sub-source domains $\mathcal{D}_{ssub_M}$,..., $\mathcal{D}_{ssub_M}$ by Eq.(6) (7)
repeat
  for $k = 1$ to M **do**
    Sample bs samples $\{x_i^{ssub_k}\}_{i=1}^{bs}$ from $\mathcal{D}_{ssub_k}$
    Sample bs samples $\{x_j^t\}_{j=1}^{bs}$ from $\mathcal{D}_T$
    Feed $x_i^{ssub_k}$ and x_j^t into F_k and get $F_k(x_i^{ssub_k})$, $F_k(x_j^t)$
    Feed $F_k(x_i^{ssub_k})$ into C_k and get $C_k(F_k(x_i^{ssub_k}))$
    Feed $F_k(x_i^{ssub_k})$, $F_k(x_j^t)$ into D_k and get $D_k(F_k(x_i^{ssub_k}))$, $D_k(F_k(x_j^t))$
    Calculate $\mathcal{L}_{\text{cls}}$ using $C_k(F_k(x_i^{ssub_k}))$ by Eq.(8)
    Calculate $\mathcal{L}_{\text{adv}}$ using $D_k(F_k(x_i^{ssub_k}))$, $D_k(F_k(x_j^t))$ by Eq.(9)
    Calculate $\mathcal{L}_{\text{total}}$ by Eq.(10)
    Update F_k, C_k, D_k to minimize $\mathcal{L}_{\text{total}}$
  end for
until convergence

where 147 is the time length and 16 is the channel number. In the intra-subject setting, classic methods like Random Forest, XGBoost, LSTM and 1D-CNN network achieved the accuracy of at least 98%. More details are in the supplement. All experiments use leave-one-subject-out cross-validation (LOSO-CV) with a common vote threshold of 50%. In this setup, each subject acts as the target domain $\mathcal{D}_T$, while the rest are the source domain $\mathcal{D}_S$ rotating through all patients.

Falling Risk Assessment Dataset (FRA) (Hu et al. 2022) It is a plantar pressure dataset comprises 48 subjects and 7,462 samples, with 23 high-risk subjects (labeled as 1) and 25 low-risk subjects (labeled as 0). The plantar pressure data has the shape $x \in \mathbb{R}^{69\times16}$, where 69 is the time length and 16 is the channel number. All experiments are also under the setting of LOSO-CV. Additionally, experiments on FRA follow a 11-step setting with a 50% threshold as (Wu et al. 2022). That indicates we split all target data into several segments, each containing 11 neighboring samples. If 50% or more of the samples in a segment are classified as high risk, the entire segment is considered high risk. Performance is calculated by segments, not subjects. This challenging setup closely mimics real-world scenarios.

Baselines and Implementation Details

Baselines We compare MSSDA with various kinds of transfer learning methods that are wildly used in biomedical signal process, including D-CORAL(Sun and Saenko 2016), DAN(Long et al. 2015), JAN(Long et al. 2017), MCD(Saito et al. 2018), DANN (Ganin et al. 2016), CDAN (Long et al. 2018) and MhNet (Wu et al. 2022). Note that ERM refers to Empirical Risk Minimization, which means to train the model without transfer learning loss.

Implementation Details All methods are implemented using the PyTorch framework and reproduced on a GeForce GTX 4060. In Stage 1, we utilize a network with 4 layers of 1D CNN for contrastive learning on DFN-DS, and a network with 3 layers of 1D CNN for the FRA. In Stage 3, the feature extractor for DFN-DS remains consistent across all methods, comprising 7 layers of 1D CNN, while the FRA uses 3 layers of 1D CNN. Additionally, the domain discriminator and classifier frameworks in the domain-invariant methods are identical, featuring 3-layer fully connected networks for DFN-DS and 2-layer fully connected networks for FRA. All experiments are conducted after balancing the dataset using data reuse techniques.

While in stage 1, we train the feature extractor F_0 by contrastive learning for 5000 epochs using AdamW as the optimizer with initial learning rate of 5e-3, a batch size of 64 for DFN-DS. As for FRA, we set the initial learning rate as 1e-3 with a batch size of 32, other paremeters remain the same. In stage 3, same as other methods, we use Adam as our optimizer with a weight decay set to 1e-4. We select the learning rate lr from {1e-2, 8e-3, 5e-3} for best performance. Additionally, we select the weights of the domain adaptation loss from $\{0.2, 0.5, 1, 2\}$ to get the best performance of the models, which is 1 for our method in FRA and 1.5 in DFN-DS. For both DFN-DS and FRA, M is set to 2.

Method	Precision	Recall	Accuracy	F1
ERM	0.767	0.596	0.593	0.671
D-CORAL	0.731	0.840	0.674	0.782
DAN	0.851	0.914	<u>0.830</u>	<u>0.882</u>
JAN	0.844	0.809	0.763	0.826
MCD	<u>0.908</u>	0.840	<u>0.830</u>	0.873
DANN	0.796	0.829	0.733	0.813
CDAN	0.767	**0.979**	0.778	0.860
MhNet	0.835	0.702	0.696	0.763
Ours	**0.916**	<u>0.926</u>	**0.889**	**0.921**

Table 1: Performance comparison of classification on DFN-DS.

Comparison Results

Results on DFN-DS As shown in Table **??**, comparing all methods on our proposed dataset DFN-DS, our approach excels across almost all metrics. In terms of recall, while our

Method	Precision	Recall	Accuracy	F1
ERM	0.491	0.717	0.696	0.583
D-CORAL	0.615	0.645	0.642	0.630
DAN	0.618	0.625	0.663	0.622
JAN	0.542	0.548	0.591	0.545
MCD	0.653	0.749	0.723	0.698
DANN	0.633	0.530	0.628	0.577
CDAN	0.583	0.647	0.707	0.613
MhNet	0.697	0.764	**0.730**	0.729
Ours	**0.710**	**0.796**	0.729	**0.750**

Table 2: Performance comparison of classification on FRA.

method ranks second, it still demonstrates substantial improvement compared to alternative approaches. Most importantly, our method achieves an accuracy that is at least 5.9% higher than other methods, which is quite remarkable. Note that the high precision of our method indicates that few patients without DFN are mistakenly predicted as having DFN. Specifically, the cluster number evaluated by BIC is 6 on DFN-DS.

Results on FRA As shown in Table **??**, under challenging and real-time conditions, our approach excels across almost all metrics. We achieve the highest recall at the cost of a 0.001% lower accuracy compared to MhNet. Note that we label gaits with high fall risk as positive cases. Thus, recall is a critical metric in our context, as it emphasizes our ability to identify individuals at high risk of falling, which is significantly more important in real-world scenarios. Specifically, the number of clusters evaluated by BIC is 11 on FRA.

Further Analysis

Ablation Study

In this section, we describe ablation studies to investigate the effect of different components of our method with $M = 2$ on DFN-DS.

SA	MA	Select	Prec	Recall	Acc	F1
			0.767	0.596	0.593	0.670
✔			0.796	0.829	0.733	0.813
	✔		0.829	**0.926**	0.815	0.874
✔		✔	0.850	0.904	0.822	0.876
	✔	✔	**0.916**	**0.926**	**0.889**	**0.921**

Table 3: Results of the ablation study on stage 3 of our method on DFN-DS.

In Table 3, we analyze the influence of aligning the distributions of each pair of source and target domains in multiple feature spaces (MA), selecting proper sub-source domains as stage 3 in proposed method (Select) and mixing the data from all sub-source domains as a source domain and applying alignment in a single feature space, actually the application of DANN (SA). It is important to note that MA and SA are mutually exclusive processes. From the results shown in Table 3, we have the following insightful observations:

- Note that the one with MA and Select is actually our proposed method. It outperforms all other methods across all metrics, demonstrating the effectiveness of each stage.
- Perform alignment in several feature spaces (MA) overcome the alignment in a single feature space (SA), whether after sub-source domain selection or not.
- The idea that aligning a source domain and target domain in a feature space (MA) helps the model to get a better performance on recall.
- The proposed idea that we should carefully select proper domains (select) to avoid negative transfer is proved, whatever perform the alignment in several feature spaces (MA) or merely in a single feature space (SA).

GMM	Statistics	CL	Cluster number	Acc
✔			9	0.770
✔	✔		23	0.815
✔		✔	7	0.785
✔	✔	✔	6	0.889

Table 4: Results of the ablation study on stage 1 and stage 2 of our method on DFN-DS.

In Table 4, we analyze the influence of convolutional feature statistics (Statistics) and contrastive learning (CL) used in stages 1 and 2 of our method on DFN-DS. All experiments include sub-source selection in stage 3. Each component contributes to the improvement of the final accuracy. Additionally, we observe the effectiveness of contrastive learning as a strategy for expanding the distances between samples. After processing with the contrastive learning module, the data structure becomes clearer and more separable, enabling GMM to identify fewer but more representative clusters.

Method	Specificity	Recall	Acc
MCDCD	0.00	**1.00**	0.696
MS-MDA	0.00	**1.00**	0.696
DSAN	0.122	0.936	0.689
MS-DANN	0.561	0.926	0.815
SFDA	0.220	0.755	0.593
MJD	0.512	0.606	0.578
Ours	**0.805**	0.926	**0.889**

Table 5: Results of the performance using different dataset partitioning method on DFN-DS.

Other Strategy to Divide Source Domain

In Table 5, we compare our method of dividing the source domain by convolutional feature statistics with other conventional MDA methods. MS-MDA (Chen et al. 2021), MCDCD (Guo, Gu, and Yang 2021), SFDA (Wu et al. 2023) and MJD (Chen 2024) split the source dataset by individuals,

while DSAN (Zhu et al. 2020) splits the source dataset by labels. MS-DANN, which corresponds to the method with MA only in Table 3, can be seen as a variant of MS-MDA that uses our proposed dataset partitioning method. We found that MCDCD and MFSAN are sensitive to the target domain with label 1 but not label 0, leading to performance drops during training. This issue persists even with balanced datasets. We believe this is due to the a patient a domain setting, which results the lack of diversity in each branch network and leads to a more extreme lack of diversity in information compared to DSAN. This also leads to a shortage of mid-domain samples to bridge the source and target domains and mitigate the significant domain gap. This highlights the effectiveness of the dataset division method proposed in this paper.

Strategy	Precision	Recall	Acc	F1
Top 1 min dis.	0.872	0.872	0.822	0.872
Top 2 min dis.(Ours)	**0.916**	0.926	**0.889**	**0.921**
Top 3 min dis.	0.862	0.904	0.837	0.885
Top 1 min sum	0.821	0.926	0.807	0.870
Top 2 min sum	0.822	0.936	0.815	0.875
Top 3 min sum	0.840	**0.947**	0.837	0.890
All in usage	0.829	0.926	0.815	0.874

Table 6: Results of the performance using different strategies to select sub-source domain(s) as source domain on DFN-DS.

Other Strategies to Select Sub-Source Domains

In Table 6, dis. stands for the distance calculated in Eq. (5) while sum refers to the total sum of Euclidean distances between all target samples and the cluster centers. All in usage refers that we use all sub-source domains to train the model, which corresponds to the method with MA only in Table 3. Considering comprehensively, our strategy is the best according to the results.

Sub-source domain selection	Specificity	Recall	Acc
mixed (DANN)	0.512	0.829	0.733
Source = $\mathcal{D}_{sub_1}$	0.512	**0.968**	0.830
Source = $\mathcal{D}_{sub_2}$	0.537	0.872	0.770
Source = $\mathcal{D}_{sub_3}$	0.585	0.947	0.837
Source = $\mathcal{D}_{sub_4}$	**0.927**	0.755	0.807
Source = $\mathcal{D}_{sub_5}$	0.561	0.947	0.830
Source = $\mathcal{D}_{sub_6}$	0.683	0.862	0.807
Source = $\mathcal{D}_{sub_1}$ and $\mathcal{D}_{sub_4}$	0.659	0.926	0.844
Top 2 min dis.(Ours)	0.805	0.926	**0.889**

Table 7: Results of the performance using different sub-source domains as source domain on DFN-DS.

Sub-source Domain Performance

In Table 7, we firstly compare the performance of using different single sub-source domain as the source domain on DANN versus using the mixed source domain on DANN with DFN-DS. The results show that expect $\mathcal{D}_{sub_4}$, all other sub-source domain outperform DANN on all metrics. $\mathcal{D}_{sub_4}$, however, exhibits significantly better performance on negative samples. These experiments reveal that not all source samples are beneficial for alignment. In a data-driven manner, we set $\mathcal{D}_{sub_1}$ and $\mathcal{D}_{sub_4}$ as the selected sub-source domains used in stage 3. However, the performance of this setup is worse than our proposed approach.

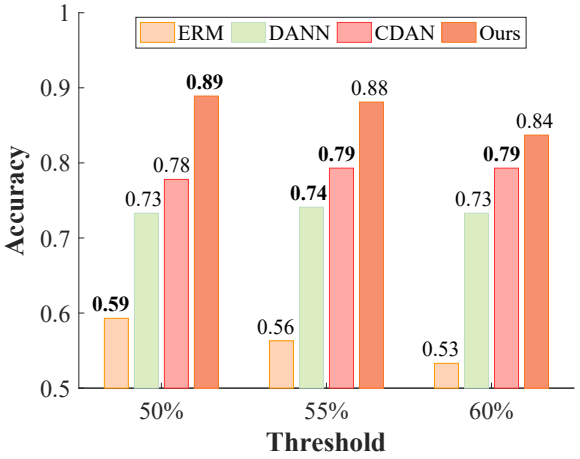

Figure 3: Results of the accuracy using other thresholds on DFN-DS.

Performance in Other Thresholds

In Figure 3, we compare the accuracy of our method with other domain-adversarial methods at different thresholds. These transfer learning methods improve accuracy on the target domain, but our method's accuracy drops more significantly as the threshold increases, likely due to excessive pursuit of transferability (Chen et al. 2019; Cui et al. 2022, 2020). This issue poses a challenge to the credibility of our approach and will be a focus of our future work.

Conlusion

We propose a dataset for Diabetic Foot Neuropathy (DFN) recognition, which includes continuous plantar pressure data from 94 DM patients with DFN and 41 DM patients without DFN. Previous works in biomedical signal processing either divide the dataset by patient or do not separate the dataset at all. Additionally, few of these studies carefully select data to avoid negative transfer. Our framework addresses these shortcomings. It divides the dataset based on convolutional feature statistics and employs a straightforward yet effective strategy to select appropriate sub-source domains for multi-source domain adaptation, simultaneously aligning the domain-specific distributions of each source-target domain pair. Extensive experiments on two plantar pressure datasets demonstrate the effectiveness of the proposed framework.

Acknowledgments

The work is supported by supported by The Taihu Lake Innovation Fund for the School of Future Technology of South China University of Technology (2024B105611003).

References

Association, A. D.; et al. 1999. Consensus development conference on diabetic foot wound care: 7-8 April 1999, Boston, MA. *Advances in Skin & Wound Care*, 12(7): 353–361.

Ben-David, S.; Blitzer, J.; Crammer, K.; Kulesza, A.; Pereira, F.; and Vaughan, J. W. 2010. A theory of learning from different domains. *Machine learning*, 79: 151–175.

Ben-David, S.; Blitzer, J.; Crammer, K.; and Pereira, F. 2006. Analysis of representations for domain adaptation. *Advances in neural information processing systems*, 19.

Botros, F. S.; Taher, M. F.; ElSayed, N. M.; and Fahmy, A. S. 2016. Prediction of diabetic foot ulceration using spatial and temporal dynamic plantar pressure. In *2016 8th Cairo international biomedical engineering conference (CIBEC)*, 43–47. IEEE.

Carls, G. S.; Gibson, T. B.; Driver, V. R.; Wrobel, J. S.; Garoufalis, M. G.; DeFrancis, R. R.; Wang, S.; Bagalman, J. E.; and Christina, J. R. 2011. The economic value of specialized lower-extremity medical care by podiatric physicians in the treatment of diabetic foot ulcers. *Journal of the American Podiatric Medical Association*, 101(2): 93–115.

Chen, H.; Jin, M.; Li, Z.; Fan, C.; Li, J.; and He, H. 2021. MS-MDA: Multisource marginal distribution adaptation for cross-subject and cross-session EEG emotion recognition. *Frontiers in Neuroscience*, 15: 778488.

Chen, S. 2024. Multi-source domain adaptation with mixture of joint distributions. *Pattern Recognition*, 149: 110295.

Chen, X.; Wang, S.; Long, M.; and Wang, J. 2019. Transferability vs. discriminability: Batch spectral penalization for adversarial domain adaptation. In *International conference on machine learning*, 1081–1090. PMLR.

Cho, N. H.; Shaw, J. E.; Karuranga, S.; Huang, Y.; da Rocha Fernandes, J. D.; Ohlrogge, A.; and Malanda, B. 2018. IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. *Diabetes research and clinical practice*, 138: 271–281.

Coppini, D.; Wellmer, A.; Weng, C.; Young, P.; Anand, P.; and Sönksen, P. 2001. The natural history of diabetic peripheral neuropathy determined by a 12 year prospective study using vibration perception thresholds. *Journal of clinical neuroscience*, 8(6): 520–524.

Cui, Q.; Zhao, B.; Chen, Z.-M.; Zhao, B.; Song, R.; Zhou, B.; Liang, J.; and Yoshie, O. 2022. Discriminability-transferability trade-off: An information-theoretic perspective. In *European Conference on Computer Vision*, 20–37. Springer.

Cui, S.; Wang, S.; Zhuo, J.; Li, L.; Huang, Q.; and Tian, Q. 2020. Towards discriminability and diversity: Batch nuclear-norm maximization under label insufficient situations. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 3941–3950.

Deng, F.; Tu, S.; and Xu, L. 2021. Multi-source unsupervised domain adaptation for ECG classification. In *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 854–859. IEEE.

DuBrava, S.; Mardekian, J.; Sadosky, A.; Bienen, E. J.; Parsons, B.; Hopps, M.; and Markman, J. 2017. Using random forest models to identify correlates of a diabetic peripheral neuropathy diagnosis from electronic health record data. *Pain Medicine*, 18(1): 107–115.

Fernando, M.; Crowther, R.; Lazzarini, P.; Sangla, K.; Cunningham, M.; Buttner, P.; and Golledge, J. 2013. Biomechanical characteristics of peripheral diabetic neuropathy: A systematic review and meta-analysis of findings from the gait cycle, muscle activity and dynamic barefoot plantar pressure. *Clinical biomechanics*, 28(8): 831–845.

Ganin, Y.; Ustinova, E.; Ajakan, H.; Germain, P.; Larochelle, H.; Laviolette, F.; March, M.; and Lempitsky, V. 2016. Domain-adversarial training of neural networks. *Journal of machine learning research*, 17(59): 1–35.

Guo, Y.; Gu, X.; and Yang, G.-Z. 2021. MCDCD: Multi-source unsupervised domain adaptation for abnormal human gait detection. *IEEE Journal of Biomedical and Health Informatics*, 25(10): 4017–4028.

Hu, G.; Jin, J.; Song, Z.; Wu, S.; Shu, L.; Xie, J.; Ou, J.; Chen, Z.; and Xu, X. 2022. A dataset for falling risk assessment of the elderly using wearable plantar pressure. In *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2900–2904. IEEE.

Kastenbauer, T.; Sauseng, S.; Sokol, G.; Auinger, M.; and Irsigler, K. 2001. A prospective study of predictors for foot ulceration in type 2 diabetes. *Journal of the American Podiatric Medical Association*, 91(7): 343–350.

Lai, J.; Tan, H.; Wang, J.; Ji, L.; Guo, J.; Han, B.; Shi, Y.; Feng, Q.; and Yang, W. 2023. Practical intelligent diagnostic algorithm for wearable 12-lead ECG via self-supervised learning on large-scale dataset. *Nature Communications*, 14(1): 3741.

Lian, X.; Qi, J.; Yuan, M.; Li, X.; Wang, M.; Li, G.; Yang, T.; and Zhong, J. 2023. Study on risk factors of diabetic peripheral neuropathy and establishment of a prediction model by machine learning. *BMC Medical Informatics and Decision Making*, 23(1): 146.

Liu, L. 2017. *Research on early diagnosis system for diabetic foot based on plantar blood pressure*. Ph.D. diss., Dept. of Biomedical Engineering, South Central University for Nationalities, Wuhan, China.

Long, M.; Cao, Y.; Wang, J.; and Jordan, M. 2015. Learning transferable features with deep adaptation networks. In *International conference on machine learning*, 97–105. PMLR.

Long, M.; Cao, Z.; Wang, J.; and Jordan, M. I. 2018. Conditional adversarial domain adaptation. *Advances in neural information processing systems*, 31.

Long, M.; Zhu, H.; Wang, J.; and Jordan, M. I. 2017. Deep transfer learning with joint adaptation networks. In *International conference on machine learning*, 2208–2217. PMLR.

Lu, W.; Chen, Y.; Wang, J.; and Qin, X. 2021. Cross-domain activity recognition via substructural optimal transport. *Neurocomputing*, 454: 65–75.

Matsuura, T.; and Harada, T. 2020. Domain generalization using a mixture of multiple latent domains. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, 11749–11756.

Mu, F.; Gu, X.; Guo, Y.; and Lo, B. 2020. Unsupervised domain adaptation for position-independent imu based gait analysis. In *2020 IEEE SENSORS*, 1–4. IEEE.

Peng, X.; Bai, Q.; Xia, X.; Huang, Z.; Saenko, K.; and Wang, B. 2019. Moment matching for multi-source domain adaptation. In *Proceedings of the IEEE/CVF international conference on computer vision*, 1406–1415.

Saito, K.; Watanabe, K.; Ushiku, Y.; and Harada, T. 2018. Maximum classifier discrepancy for unsupervised domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3723–3732.

Sazonov, E. S.; Hegde, N.; and Tang, W. 2013. Development of SmartStep: An insole-based physical activity monitor. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 7209–7212. IEEE.

Schwarz, G. 1978. Estimating the dimension of a model. *The annals of statistics*, 461–464.

She, Q.; Zhang, C.; Fang, F.; Ma, Y.; and Zhang, Y. 2023. Multisource associate domain adaptation for cross-subject and cross-session EEG emotion recognition. *IEEE Transactions on Instrumentation and Measurement*, 72: 1–12.

Sun, B.; and Saenko, K. 2016. Deep coral: Correlation alignment for deep domain adaptation. In *Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14*, 443–450. Springer.

van Schie, C. H. 2008. Neuropathy: mobility and quality of life. *Diabetes/metabolism research and reviews*, 24(S1): S45–S51.

Wafai, L.; Zayegh, A.; Woulfe, J.; and Begg, R. 2014. Automated classification of plantar pressure asymmetry during pathological gait using artificial neural network. In *2nd Middle East Conference on Biomedical Engineering*, 220–223. IEEE.

Wang, D.; Ouyang, J.; Zhou, P.; Yan, J.; Shu, L.; and Xu, X. 2020. A novel low-cost wireless footwear system for monitoring diabetic foot patients. *IEEE Transactions on Biomedical Circuits and Systems*, 15(1): 43–54.

Wang, Y.; Han, Y.; Wang, H.; and Zhang, X. 2024. Contrast everything: A hierarchical contrastive framework for medical time-series. *Advances in Neural Information Processing Systems*, 36.

Wu, S.; Ou, J.; Shu, L.; Hu, G.; Song, Z.; Xu, X.; and Chen, Z. 2022. MhNet: Multi-scale spatio-temporal hierarchical network for real-time wearable fall risk assessment of the elderly. *Computers in biology and medicine*, 144: 105355.

Wu, S.; Shu, L.; Song, Z.; and Xu, X. 2023. SFDA: domain adaptation with source subject fusion based on multi-source and single-target fall risk assessment. *IEEE transactions on neural systems and rehabilitation engineering*.

Xie, Y.; Lin, Y.; Cai, W.; Xu, X.; Zhang, H.; Du, Y.; and He, S. 2024. D3still: Decoupled Differential Distillation for Asymmetric Image Retrieval. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 17181–17190.

Xie, Y.; Zhang, H.; Xu, X.; Zhu, J.; and He, S. 2023. Towards a smaller student: Capacity dynamic distillation for efficient image retrieval. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 16006–16015. IEEE.

Zhu, Y.; Zhuang, F.; and Wang, D. 2019. Aligning domain-specific distribution and classifier for cross-domain classification from multiple sources. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, 5989–5996.

Zhu, Y.; Zhuang, F.; Wang, J.; Ke, G.; Chen, J.; Bian, J.; Xiong, H.; and He, Q. 2020. Deep subdomain adaptation network for image classification. *IEEE transactions on neural networks and learning systems*, 32(4): 1713–1722.