scientific reports

Check for updates

OPEN

Correlation enhanced distribution adaptation for prediction of fall risk

Ziqi Guo[1], Teresa Wu[2], Thurmon E. Lockhart[3], Rahul Soangra[4] & Hyunsoo Yoon[5]✉

With technological advancements in diagnostic imaging, smart sensing, and wearables, a multitude of heterogeneous sources or modalities are available to proactively monitor the health of the elderly. Due to the increasing risks of falls among older adults, an early diagnosis tool is crucial to prevent future falls. However, during the early stage of diagnosis, there is often limited or no labeled data (expert-confirmed diagnostic information) available in the target domain (new cohort) to determine the proper treatment for older adults. Instead, there are multiple related but non-identical domain data with labels from the existing cohort or different institutions. Integrating different data sources with labeled and unlabeled samples to predict a patient's condition poses a significant challenge. Traditional machine learning models assume that data for new patients follow a similar distribution. If the data does not satisfy this assumption, the trained models do not achieve the expected accuracy, leading to potential misdiagnosing risks. To address this issue, we utilize domain adaptation (DA) techniques, which employ labeled data from one or more related source domains. These DA techniques promise to tackle discrepancies in multiple data sources and achieve a robust diagnosis for new patients. In our research, we have developed an unsupervised DA model to align two domains by creating a domain-invariant feature representation. Subsequently, we have built a robust fall-risk prediction model based on these new feature representations. The results from simulation studies and real-world applications demonstrate that our proposed approach outperforms existing models.

Keywords Unsupervised domain adaptation, Machine learning, Classification, Fall risk

Vast volumes of unlabeled data are generated and made available in numerous domains. In the context of machine learning, a domain refers to a subset of the larger data space that is relevant for a specific task or application. However, acquiring sufficient labeled data can be exceedingly costly and sometimes impractical. For example, on average, each pixel-level image in the Cityscapes dataset required 1.5 h to complete the annotation[1]. Domain adaptation (DA) addresses the limited labeled data issue by aligning two distinct datasets: one from a source domain and the other from a target domain. The source domain contains a large amount of labeled data on which classifiers can be reliably trained. The target domain broadly refers to a dataset assumed to have different characteristics from the source domain, where those classifiers are applied. Several example scenarios require domain adaptation (DA). In computer vision tasks, objects might come from multiple sources, each with different backgrounds, object styles, and locations[2–6]. In activity recognition tasks, sensors might be placed in different body locations[7,8]. In speech recognition, voices may come from different speakers[9,10]. In sentiment analysis, various text sources, such as electronics or DVDs, are used for analysis[11,12]. In healthcare, acquiring labeled data and large samples is even more challenging. For instance, in medical image analysis, the major challenge in constructing reliable and robust models is the lack of labeled data[13,14]. Clinical outcomes might be sourced from different machines and healthcare providers. Variations between different data sources can significantly reduce prediction accuracy. These problems are studied in DA, where the model is learned on one dataset (i.e., source domain) and then transferred to a target dataset (i.e., target domain) with different distribution properties.

Although machine learning approaches for supervised learning have performed well, they assume that training and testing data are drawn from the same distribution, which may not always be true. To complement this challenge, DA aims to align the target to the source domain by creating a domain-invariant feature representation. After adaptation, it becomes a standard machine learning problem that assumes test data are drawn from a similar distribution as the training data. In this paper, we propose an unsupervised DA method that specifically

[1]Department of Systems Science and Industrial Engineering, The State University of New York at Binghamton, Binghamton, USA. [2]School of Computing and Augmented Intelligence, Arizona State University, Tempe, USA. [3]School of Biological and Health Systems Engineering, Arizona State University, Tempe, USA. [4]Department of Physical Therapy, Chapman University, Orange, USA. [5]Department of Industrial Engineering, Yonsei University, Seoul, Korea. ✉email: hs.yoon@yonsei.ac.kr

addresses situations where labeled data are available only in the source domain, and the target domain is unlabeled, which is common in practice.

According to a literature review[15], existing DA methods can be organized into two categories: (a) feature transformation and (b) instance weighting. Feature transformation either performs feature space alignment by exploring the subspace geometrical structure, such as subspace alignment (SA)[16], CORrelation ALignment (CORAL)[17], and geodesic flow kernel (GFK)[5], or distribution adaptation to reduce the distribution divergence between domains, such as transfer component analysis (TCA)[18] and joint distribution adaptation (JDA)[19]. Instance reweighting reweights the samples from the source domain to the target based on the weighting methods[20,21]. The challenge with existing methods is degenerated feature transformation[22], where both subspace alignment and distribution adaptation can reduce the divergence between domains but not eliminate it. Subspace alignment only considers the subspace or manifold structure, failing to achieve complete feature alignment. Conversely, distribution adaptation reduces the distribution distance in the original feature space but often distorts features, making it more challenging to reduce the divergence. Therefore, exploiting both the advantages of subspace alignment and distribution adaptation is significant for further developing DA. This study proposes a novel DA method to address this challenge.

Unsupervised learning assumes the availability of labeled source data and unlabeled target data. Several unsupervised domain adaptation (DA) methods are described in a literature review[23]. Domain-invariant feature learning methods aim to align the source and target domains by creating a domain-invariant feature representation, where features follow the same distribution regardless of the input's source or target domain. Typically, this is achieved through a feature extractor neural network[17,24–26]. Domain mapping methods, on the other hand, use adversarial techniques to create a pixel-level map from one domain to another, often accomplished with a conditional GAN[27–29]. Normalization statistics methods leverage normalization layers like batch normalization commonly found in neural networks[30,31]. Existing unsupervised DA methods predominantly emphasize neural network-based approaches, but they may perform poorly in cases with a small sample size and a limited number of features. This can be attributed to the fact that neural networks typically require large amounts of data to learn meaningful representations and can suffer from overfitting when the number of features is limited. Therefore, to address this shortcoming, we propose our shallow unsupervised DA approach, Correlation Enhanced Distribution Adaptation (CEDA).

Domain adaptation has garnered considerable attention in healthcare applications in recent years, particularly in computer-aided medical image analysis[32–34], due to its ability to reuse pre-trained models from related domains. Many other healthcare problems also face the challenge of lacking labeled data. This study extends the application of domain adaptation, especially unsupervised DA, to sensor-based prognosis.

Of particular interest in this research is fall detection. Falls pose significant threats to the health of older adults and can hinder their ability to remain independent. As CDC reports suggest, 3 million older people are treated in emergency departments for fall injuries each year, and fall death rates in the U.S. increased by 30% from 2007 to 2016. Therefore, fall prevention is a critical component of healthcare for the senior community. In the realm of fall risk assessment, particularly for older adults, there is a recognized importance of both intrinsic and extrinsic factors. Intrinsic factors include muscle strength[35], balance[36], and gait stability[37], whereas extrinsic factors involve elements like home hazards and footwear choices[38]. Recently, wearable sensors have become invaluable in assessing fall risk, especially through the use of accelerometers and gyroscopes to capture a variety of movement characteristics. Diverse feature sets have been explored in fall risk assessment, including nonlinear dynamics. Measures such as Shannon entropy and frequency analysis, which reflect gait dynamics, have shown significantly higher values in individuals prone to falls, indicating their potential as fall risk predictors[39]. Nonlinear metrics, like multiscale entropy (MSE) and recurrence quantification analysis (RQA) applied to trunk accelerations, have demonstrated positive correlations with fall histories, suggesting their utility in identifying individuals at higher risk[40]. Koshmak et al. employed supervised feature learning to estimate fall risk probabilities, underscoring the critical importance of feature selection in effective assessment[41]. Additionally, research has highlighted the significance of integrating gait and posture analysis for enhanced precision in predicting fall risks[42]. Recent studies collectively emphasize the substantial potential of wearable sensors in delineating fall risk, particularly through examining features like entropy, complexity, multiscale entropy, and fractal properties[43–45].

This study proposes a novel approach for fall prediction using the 10-m walking test. We focus on the challenge where the fall information for the target group is unknown, while it is known for the other group. As they are different groups of people, their characteristic distributions (marginal and conditional) differ. Hence, directly using data from one group to train the classification models would not provide accurate predictions for the other group.

Methods

Formulation

Without loss of generality, we describe our method by taking a binary classification problem as the running example. The proposed formula can be directly applicable to multi-class classification problems. Assume source-domain training examples $D_S = \{\overrightarrow{x_i}\}$, $\overrightarrow{x} \in \mathbb{R}^D$ with labels $L_s = \{y_i\}$, $y \in \{1, , \ldots, L\}$, and target data $D_T = \{\overrightarrow{u_i}\}$, $\overrightarrow{u} \in \mathbb{R}^d$. Both $\overrightarrow{x}$ and $\overrightarrow{u}$ are the d-dimensional feature representations $\phi(I)$ of input I.

Proposed method

We propose the Correlation Enhanced Distribution Adaptation (CEDA) model, which combines and improves upon the CORrelation ALignment (CORAL) and Joint Distribution Adaptation (JDA) approaches, outperforming each of these methods individually. In the following section, we will provide a brief introduction to these two approaches: CORrelation ALignment (CORAL) and Joint Distribution Adaptation (JDA).

(1) CORrelation ALignment (CORAL)[17] transforms the source features to the target space by aligning the second-order statistic, the covariance. The covariances differ in the original source and target domain distributions. The researchers propose conducting source decorrelation to remove the feature correlation of the source domain and then constructing target re-correlation by adding the correlation of target features to the source domain. After these two steps, the two distributions are well aligned, and the classifiers trained on the adjusted source domain work well in the target. However, this method aligns the source distributions as a whole to the target domain, neglecting the significance of individual samples.

(2) Joint Distribution Adaptation (JDA)[19] aims to find a feature transformation that jointly minimizes the difference in marginal and conditional distributions between domains. Although no labeled data exists in the target domain, this method generates pseudo-target labels by applying a classifier f trained on the adapted labeled source to the unlabeled target. Iterative label refinement is used to improve the classifier and labeling quality. However, it has limitations in generating accurate pseudo labels for the target domain."

Our proposed method begins by employing CORAL as the first step for source decorrelation, which involves removing the feature correlation of the source domain and adding the correlation of the target to the source domain. This integrated adaptation aims to roughly align the source samples to the target domain. However, due to the presence of distribution noise, some samples may not be correctly aligned, leading to suboptimal results. To ensure accurate alignment for all samples, a further meticulous adaptation is performed. In the second step of our proposed method, we apply Joint Distribution Adaptation (JDA) to the adjusted source samples obtained from the first step. JDA has a limitation of generating pseudo-target labels in the first iteration, which can result in an inappropriate adjustment in the conditional distribution. To overcome this challenge, we utilize CORAL to provide an initial adjusted source sample for JDA. The transformed target samples are then classified using a 1-Nearest Neighbor (1NN) classifier, trained with the transformed new source samples.

Moreover, CORAL serves as a nonparametric model that does not require any parameter tuning, making it highly advantageous for unsupervised learning. It aligns the distribution of source and target features in an unsupervised manner. In our approach, CORAL transforms the source feature $\mathbf{X}_S$ to the target space $\mathbf{X}_\mathrm{T}$ by aligning the second-order statistic, the covariance. After obtaining new $\mathbf{X}_S$ by multiplying the CORAL adaptation matrix (A_CORAL) with $\mathbf{X}_S$, we train a standard classifier f (nearest neighbor in our case) on the new $\mathbf{X}_S$ to generate the initial pseudo-target labels $\widehat{\boldsymbol{y}}_T$ for the target. Subsequently, we build an MMD (Maximum Mean Discrepancy) matrix $\mathbf{M}$ (Gretton et al., 2008):

$$(M_0)_{ij} = \begin{cases} \frac{1}{n_s n_s}, & x_i, x_j \in \mathcal{D}_s \\ \frac{1}{n_t n_t}, & x_i, x_j \in \mathcal{D}_t \\ \frac{-1}{n_s n_t}, & otherwise \end{cases} \tag{1}$$

which is adopted as the distance measurement for the objective of reducing the difference between marginal distributions $P_s(\boldsymbol{X}_s)$ and $P_t(\boldsymbol{X}_T)$. An MMD matrix $\{\mathbf{M}_C\}_{c=1}^{C}$ is then constructed based on class labels, used as the distance measurement for minimizing the difference between conditional distribution, as follows:

$$(M_c)_{ij} = \begin{cases} \frac{1}{n_s^{(c)} n_s^{(c)}}, x_i, x_j \in \mathcal{D}_s^{(c)} \\ \frac{1}{n_t^{(c)} n_t^{(c)}}, x_i, x_j \in \mathcal{D}_t^{(c)} \\ \frac{-1}{n_s^{(c)} n_t^{(c)}}, \begin{cases} x_i \in \mathcal{D}_s^{(c)}, x_j \in \mathcal{D}_t^{(c)} \\ x_j \in \mathcal{D}_s^{(c)}, x_i \in \mathcal{D}_t^{(c)} \end{cases} \\ 0, otherwise \end{cases} \tag{2}$$

Next, the optimal adaptation matrix A is calculated by solving Eq. (3) for the k smallest eigenvectors, and $\boldsymbol{Z} := \boldsymbol{A}^T \boldsymbol{X}$:

$$(\mathbf{X}\sum\nolimits_{c=0}^{C}\mathbf{M}_c\mathbf{X}^\mathrm{T} + \lambda\mathbf{I})\mathbf{A} = \mathbf{X}\mathbf{H}\mathbf{X}^\mathrm{T}\mathbf{A}\boldsymbol{\Phi} \tag{3}$$

A standard classifier $\not$ is trained on $(\mathbf{A}_S^\mathrm{T}\mathbf{X}_s, \boldsymbol{y}_S)$ to generate $\widehat{\boldsymbol{y}}_T := \not\ (\mathbf{A}_T^T \boldsymbol{X}_T)$. If we use this labeling $\widehat{\boldsymbol{y}}_T$ as the pseudo-target labels and run JDA iteratively, we can alternate improving the labeling quality until convergence. The model will return adaptation matrix $\mathbf{A}$, embedding $\mathbf{Z}$, adaptive classifier f, with the input of source data $\mathbf{X}_\mathrm{S}$, $\mathbf{y}_\mathrm{s}$, target Data $\mathbf{X}_\mathrm{T}$; #subspace bases k, regularization parameter λ.

The algorithm is summarized in the following pseudo-code:

Input: Source Data $\mathbf{X}_S$, $\mathbf{y}_s$, Target Data $\mathbf{X}_T$; #subspace bases k, regularization parameter λ.

Output: Adaptation matrix $\mathbf{A}$, embedding $\mathbf{Z}$, adaptive classifier f

begin

$C_S = cov(\boldsymbol{X}_S) + eye(size(\boldsymbol{X}_S,2))$

$C_T = cov(\boldsymbol{X}_T) + eye(size(\boldsymbol{X}_T,2))$

$\mathbf{A}_{CORAL} = C_S^{-\frac{1}{2}} * C_T^{\frac{1}{2}}$

The new adjusted source features: $\mathbf{X}_{\mathrm{Sadj}} = \mathbf{X}_S\mathbf{A}_{\mathrm{CORAL}}$

Train a standard classifier f on $(\mathbf{X}_{\mathrm{Sadj}}, \boldsymbol{y}_S)$ to generate initial pseudo-target labels $\hat{\boldsymbol{y}}_T$:=ʃ ($\boldsymbol{X}_T$)

Construct MMD matrix $\mathbf{M}_0$ by Eq. (1)

repeat

Construct MMD matrix $\{\mathbf{M}_C\}_{c=1}^{C}$ by Equation (2).

Solve the generalized eigen decomposition problem in Equation (3) and select the k smallest eigenvectors to construct the adaptation matrix $\boldsymbol{A}$, and $\boldsymbol{Z}$:=$\boldsymbol{A}^T\boldsymbol{X}$.

Train a standard classifier ʃ on $(\mathbf{A}_S^{\mathrm{T}}\mathbf{X}_{\mathrm{Sadj}}, \boldsymbol{y}_S)$ to update pseudo-target labels $\hat{\boldsymbol{y}}_T$:=ʃ ($\mathbf{A}_T^T\boldsymbol{X}_T$)

until Convergence

Return an adaptive classifier ʃ trained on $(\mathbf{A}_S^{\mathrm{T}}\mathbf{X}_{\mathrm{Sadj}}, \boldsymbol{y}_S)$

Algorithm. CEDA for unsupervised DA.

Simulation study

This section uses simulation data to demonstrate the proposed method's performance under several scenarios. The simulation data are generated as follows: the source and target domain data are sampled from a multi-dimensional normal distribution with randomly selected parameter setting. We consider a binary classification. In the source domain, the simulation data $\boldsymbol{X_s} \sim \mathcal{N}(\boldsymbol{\mu_s}, \boldsymbol{\Sigma_s})$ with corresponding responses $\boldsymbol{Y_s} \in \{0, 1\}$ and $\boldsymbol{X_t} \sim \mathcal{N}(\boldsymbol{\mu_t}, \boldsymbol{\Sigma_t})$, $\boldsymbol{Y_t} \in \{0, 1\}$ for the target domain.

Impact of sample size on model performance

In the simulation setup, while maintaining the sample mean and covariance values, change the number of samples in each class. Each dataset is constructed by randomly selecting parameter values within predefined ranges. Specifically, the mean vector μ is randomly drawn from a uniform distribution within the interval[2,5] for red class and[4,9] for blue class, across each dimension. Similarly, the covariance matrix Σ is generated by first randomly selecting diagonal elements from a uniform distribution within the range[1,3] for source samples[4,6], for target, and then applying a random orthogonal transformation to introduce off-diagonal covariance components. The dimension for each class is the same and is randomly selected from a uniform distribution within the interval[2,20]. The scatter plots of the sample distributions and the classification accuracies are illustrated in Fig. 1.

Impact of overlap between classes on model performance

We test the effects of overlap between two classes on the classification accuracies of each model by changing the mean and covariance and maintaining the number of samples at 100. In the experiment setup for this case, we use the fixed set of parameters for normal distribution.

For source: $\boldsymbol{\mu}_1 = \begin{bmatrix} 2.5 \\ 7.5 \end{bmatrix}$ and $\boldsymbol{\Sigma}_1 = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}, \boldsymbol{\mu}_2 = \begin{bmatrix} 7 \\ 4 \end{bmatrix}$ and $\boldsymbol{\Sigma}_2 = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$,

For target:

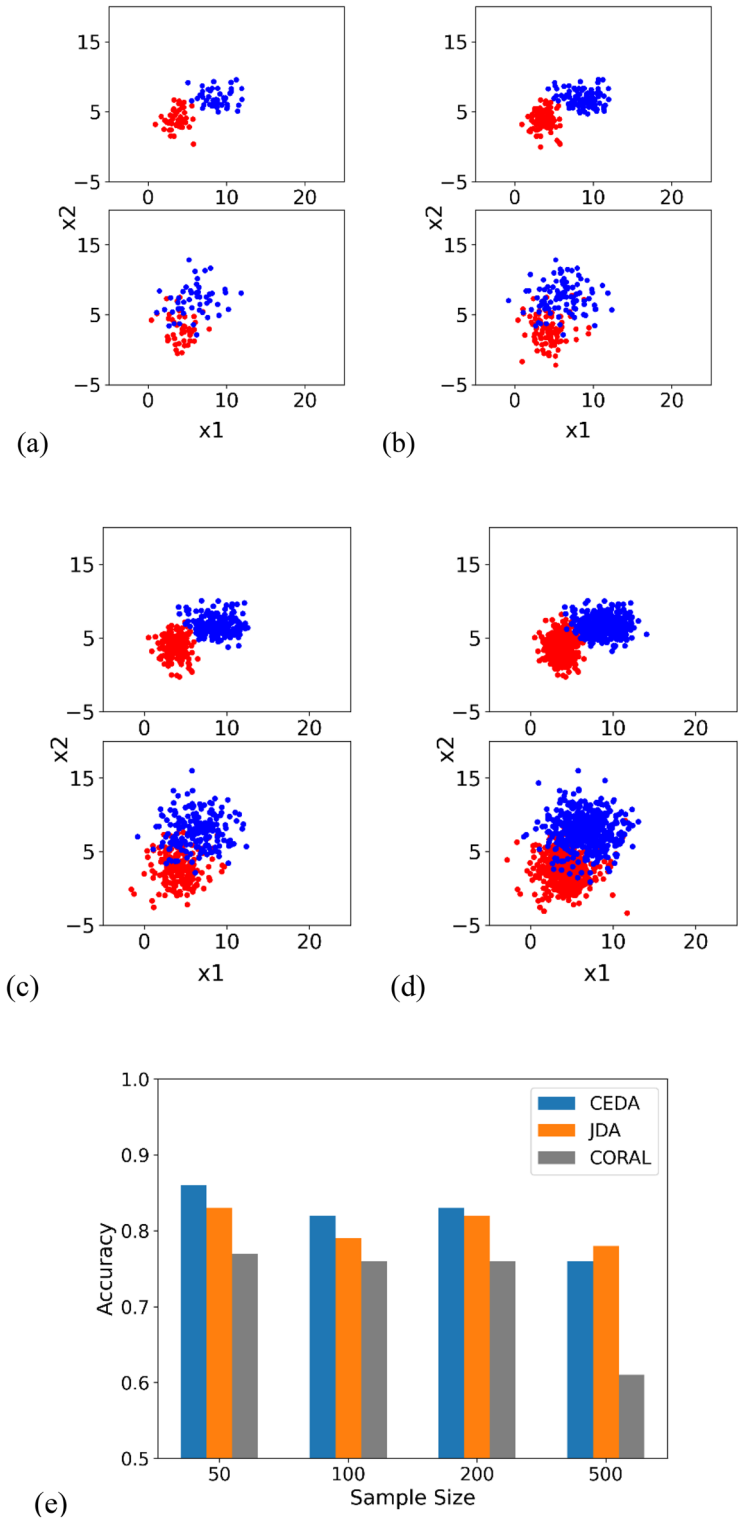

Figure 1. Scatter plots of source samples (in upper plots) and target samples (in lower plots). We visualize the first and second dimension. Two colors (red and blue) represent two classes. (**a–d**) Have 50, 100, 200, and 500 samples, respectively. (**e**) Denotes the classification accuracies at different sample sizes.

(a) $\boldsymbol{\mu}_1 = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$ and $\boldsymbol{\Sigma}_1 = \begin{bmatrix} 8 & 0 \\ 0 & 2 \end{bmatrix}, \boldsymbol{\mu}_2 = \begin{bmatrix} 13 \\ 0 \end{bmatrix}$ and $\boldsymbol{\Sigma}_2 = \begin{bmatrix} 6 & 0 \\ 0 & 1 \end{bmatrix}$,

(b) $\boldsymbol{\mu}_1 = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$ and $\boldsymbol{\Sigma}_1 = \begin{bmatrix} 8 & 0 \\ 0 & 2 \end{bmatrix}, \boldsymbol{\mu}_2 = \begin{bmatrix} 8 \\ 1 \end{bmatrix}$ and $\boldsymbol{\Sigma}_2 = \begin{bmatrix} 6 & 0 \\ 0 & 1 \end{bmatrix}$,

(c) $\boldsymbol{\mu}_1 = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$ and $\boldsymbol{\Sigma}_1 = \begin{bmatrix} 8 & 0 \\ 0 & 2 \end{bmatrix}, \boldsymbol{\mu}_2 = \begin{bmatrix} 8 \\ 2 \end{bmatrix}$ and $\boldsymbol{\Sigma}_2 = \begin{bmatrix} 6 & 0 \\ 0 & 2 \end{bmatrix}$,

(d) $\boldsymbol{\mu}_1 = \begin{bmatrix} 2.5 \\ 6 \end{bmatrix}$ and $\boldsymbol{\Sigma}_1 = \begin{bmatrix} 8 & 0 \\ 0 & 2 \end{bmatrix}, \boldsymbol{\mu}_2 = \begin{bmatrix} 7 \\ 4 \end{bmatrix}$ and $\boldsymbol{\Sigma}_2 = \begin{bmatrix} 6 & 0 \\ 0 & 2 \end{bmatrix}$,

The scatter plots of sample distributions and the classification accuracies are illustrated in Fig. 2.

Impact of noise on model performance

In this simulation study, the effect of noise on the classification accuracy of each model is tested. The mean vector μ, covariance matrix Σ, and dimension n are generated as described in "Impact of sample size on model performance". We generate 100 samples for each class, with noise added to each sample.

The noises ϵ are sampled from a uniform distribution, $\mathcal{U}_{[a,b]}$

(a) $\mathcal{E} \in [-1, 1]$
(b) $\mathcal{E} \in [-2, 2]$
(c) $\mathcal{E} \in [-3, 3]$
(d) $\mathcal{E} \in [-4, 4]$

The scatter plot in Fig. 3 illustrates the sample distribution, and the classification results.

Summary of three experiments

In the three experiments, we tested the robustness of our proposed model by (1) increasing the number of samples in each class, (2) increasing the level of overlap between the two classes, and (3) increasing the noise within each class. The results indicate that our method achieves the highest accuracies compared to JDA and CORAL under the majority of scenarios. The marginal or inferior performance of the proposed method in Figs. 1 and 2 is primarily due to the challenging nature of the datasets under certain conditions, such as significant class overlap. These scenarios are notoriously difficult for most DA methods, and our results reflect these inherent challenges.

Application in fall risk prediction

In this section, we demonstrate the application of the proposed model to predict fall risk using the dataset obtained from[46]. The human subject experimental procedures followed the principles outlined in the Declaration of Helsinki and gained approval from the Institutional Review Board (IRB) at Virginia Tech (VT), (with assigned protocol codes 11-1088 and study approval date as 10-04-2013). The research took place across four distinct community centers in Northern Virginia—Dale City, Woodbridge, Leesburg, and Manassas. The study employed consistent equipment, specifically Inertial Measurement Units (IMUs), on various days. All research activities were performed in accordance to VT-IRB regulations and guidelines and all participants provided written consent before beginning the study. Participants wear a wearable measurement device and perform a 10-m walking test, from which we extract 50 features related to linear and nonlinear gait parameters for fall risk prediction in two cohorts. The first cohort comprises 171 community-dwelling older adults with known fall information within the last six months. The second cohort consists of 49 osteoporosis patients. All participants underwent the same 10-m walking test following the same guidelines. The challenge is to accurately predict the fall risks of each individual in one group while transferring knowledge from the other group of new patients.

Data preprocessing

The dataset comprises 50 features, including 28 linear features (e.g., average step time and walking velocity) and 22 nonlinear features (e.g., anterior–posterior-signal root mean square and vertical-signal maximum line from recurrence quantification analysis). The feature correlations are identical in the two data sources. The feature correlation heatmap (Fig. 4) reveals several highly correlated features. To address potential issues with unstable predictive models and cope with small sample size problems, feature selection and dimension reduction are necessary before applying DA.

Feature selection dimension reduction

(1) Principal components analysis (PCA)[47]

PCA is a widely used technique for dimension reduction by projecting sample points onto the first few principal components (PCs) to obtain lower-dimensional data while preserving as much variation as possible. In this case study, we calculate 10 PCs from the 28 linear features and 12 PCs from the 22 nonlinear features, and then combine them into 22 PCs. This approach helps minimize the correlation between features within each category of linear and nonlinear features.

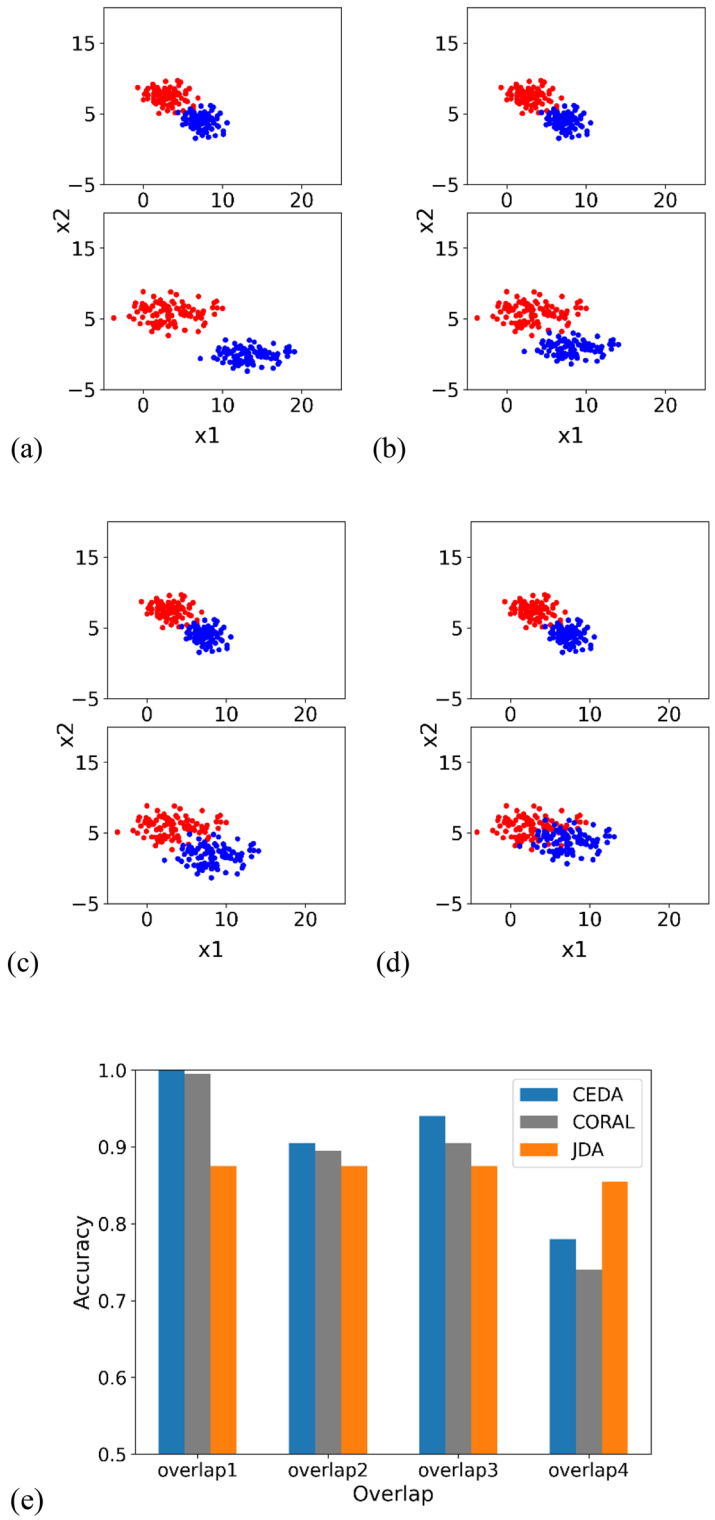

Figure 2. Scatter plots of source samples (upper plot) and target samples (lower plot). Two colors (red and blue) represent two classes. (**a–d**) Depict increasing overlap between classes. (**e**) Denotes classification accuracies at different amounts of overlap.

(2) Filter features based on mutual information[48]

Mutual information measures the mutual dependence between two variables by quantifying the "amount of information" shared between them. It is equal to zero if and only if two random variables are independ-

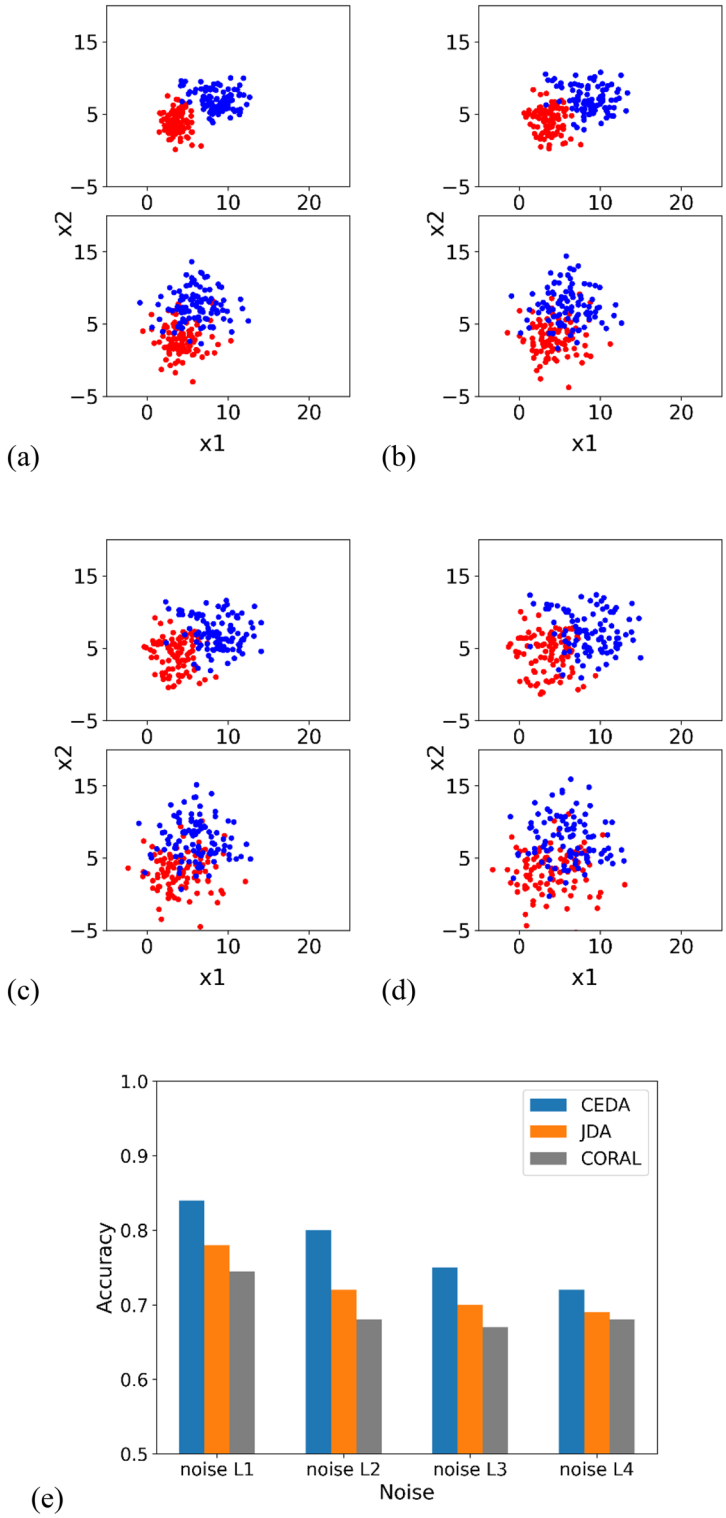

Figure 3. Scatter plots of source samples (upper plot) and target samples (lower plot). We visualize the first and second dimension. Two colors (red and blue) represent two classes. (**a–d**) Illustrate class samples with increasing noise. (**e**) Denotes classification accuracies at different amounts of overlap.

ent, with higher values indicating a higher dependency. We select the top 10 features from the original set of 50 features based on mutual information.

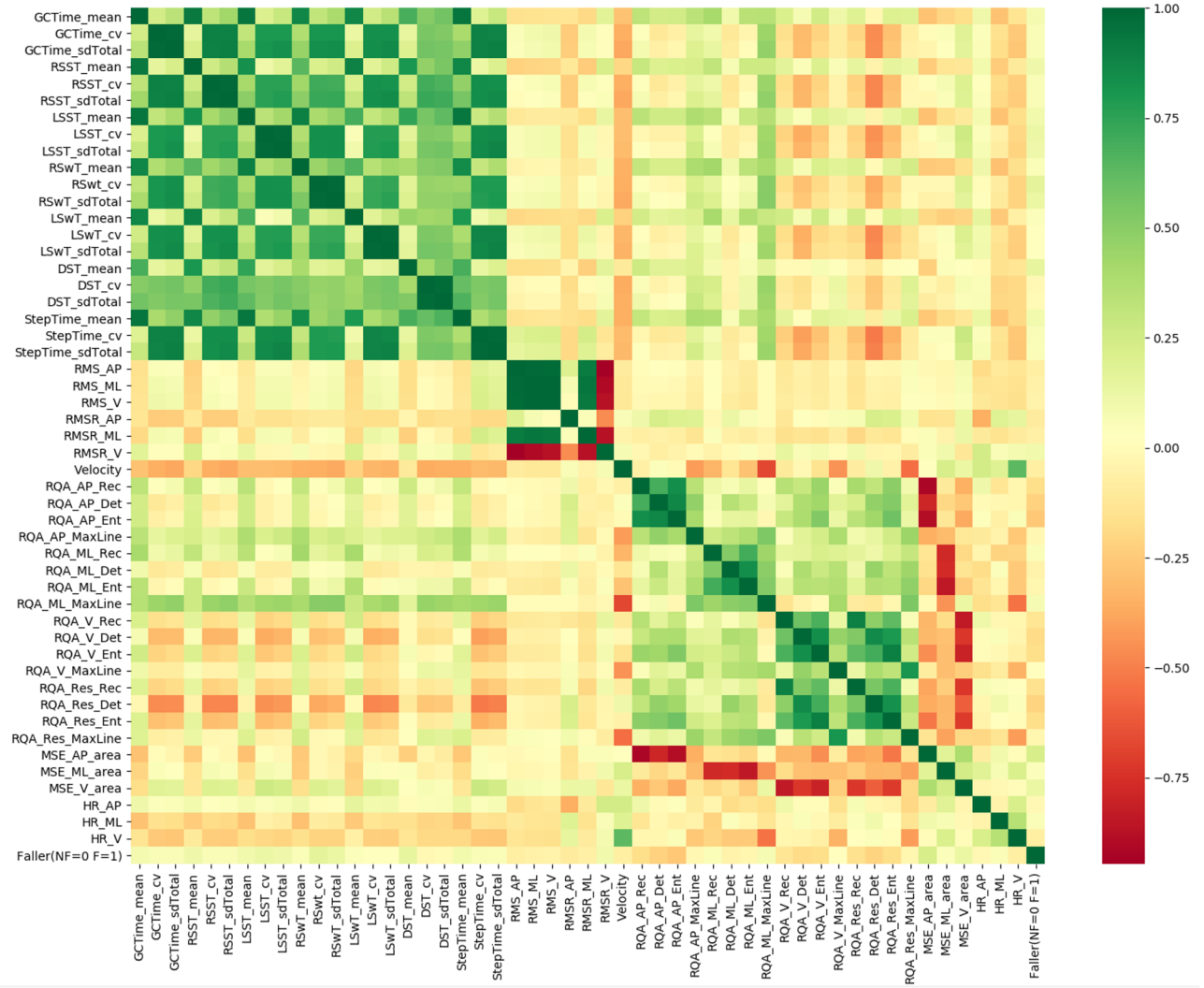

Figure 4. Heatmap of feature correlations.

Experiment results

The statistics of the two domains also illustrate that the two data sources have different characteristics of features, shown in Fig. 5. Therefore, we must adapt them for better use. Table 1 presents the classification results of directly applying models trained on the source domain to the target domain. We utilized seven classic classification models: support vector machine (SVM)[2], logistic regression (LR)[49], decision tree (DT)[50], k-nearest neighbors (KNN)[51], random forest (RF)[52], gradient boosting machine (GBM)[53] and extreme gradient boost (XGBoost)[54]. To minimize bias caused by a single method, we calculated the average of five classification accuracies.

The experiments were conducted as follows: First, we performed a stratified train and test split on the source samples (171 samples) in an 80%:20% proportion. To address the imbalance in the training data, we applied the synthetic minority over-sampling technique (SMOTE)[6] and random under-sampling technique for resampling the training set. Next, we used cross-validation to tune the optimal parameters in the classifiers. The classification model with the best parameter setting was trained on the training set and used to predict the labels for both the training and testing sets. Subsequently, we applied the model trained on the source dataset to the target samples. We conducted 15 experimental trials with different train-test splits and calculated the average accuracies as the performance measurement. The results showed that the average testing accuracy decreased from 0.7 to 0.56, indicating that directly applying the trained model from the source domain does not yield satisfactory results for the target domain.

In accordance with[19], we utilize the 1-Nearest Neighbor Classifier (1NN) as the classifier for a fair and straightforward comparison between the proposed method and baseline methods. Since the labeled source and unlabeled target data are sampled from different distributions, tuning parameters using cross-validation is not feasible. Thus, we evaluate all methods by empirically searching the parameter space to find the optimal settings and report the average results for each method. For JDA and CEDA, we search for the number of bases (k) within the range [2, 3, 4, …, 10] and the regularization parameter (λ) from the set {0.01, 0.1, 1, 10, 100}. For GFK, the parameter dimension (d) is used in the range between 1 to half of the feature dimensions, e.g. for 10 features case, d is within[1–5]. CORAL and EasyTL[55] are parametric-free methods, therefore, no parameter tuning

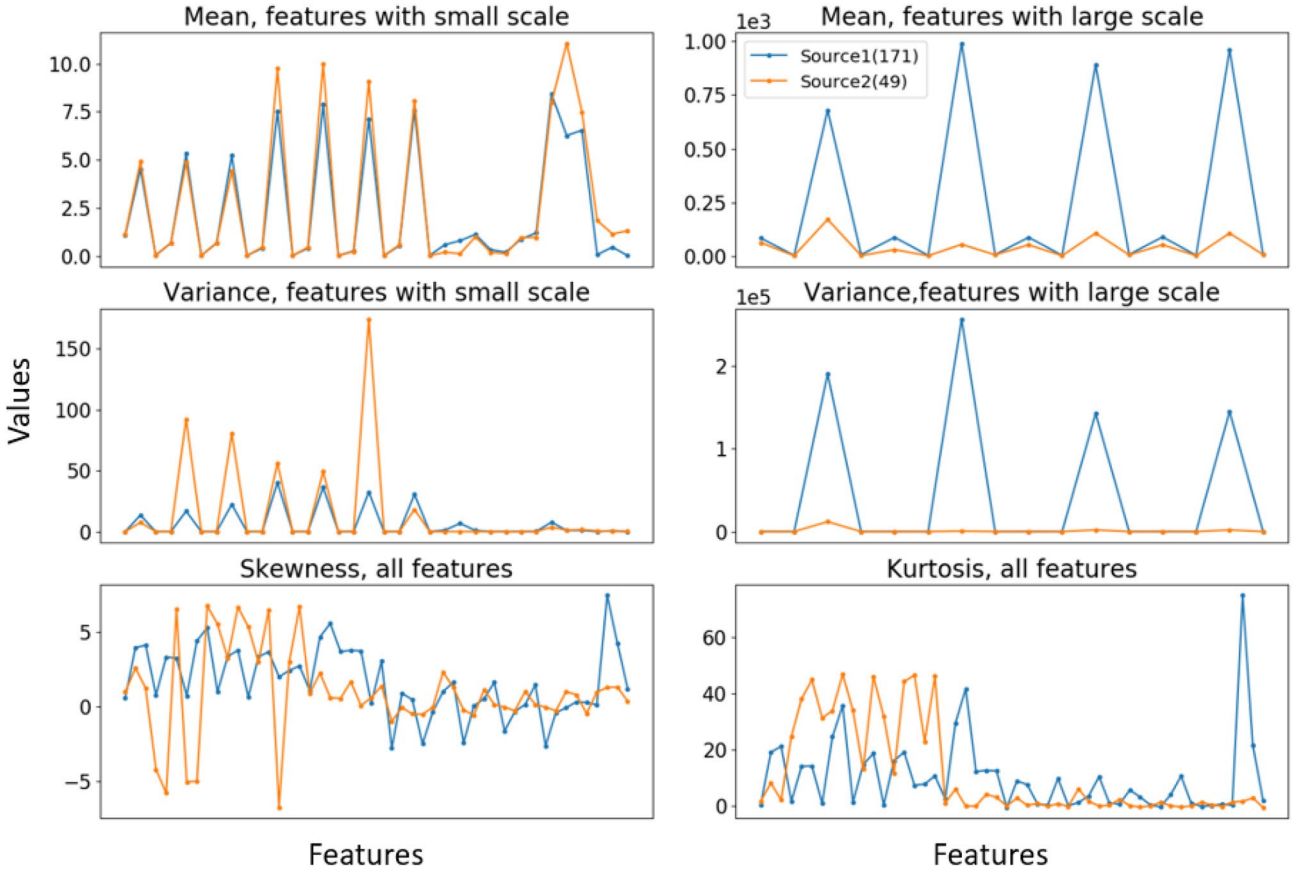

Figure 5. Mean, variance, skewness, and kurtosis of 50 features in two data sources.

	Classification accuracy based on source		Training on source testing on target	
Classifiers	Train acc	Test acc	Train acc	Test acc
SVM	0.95	0.69	0.97	0.47
LR	0.64	0.75	0.79	0.66
DT	0.93	0.69	0.91	0.57
KNN	1.00	0.65	1.00	0.67
RF	0.96	0.72	0.96	0.43
GBM	1.00	0.72	0.72	0.55
XGBoost	0.99	0.71	0.72	0.58
Average	0.92	0.70	0.87	0.56

Table 1. Classification accuracies based on source data and accuracies of directly applying models trained on source and target domains.

is needed. The experiments are conducted with different data splits five times, and we report the average accuracy along with the standard deviation.

To ensure a fair comparison and avoid data imbalances, we carefully select samples for the dataset cases: dataset 1 (source dataset) to dataset 2 (target dataset) in a ratio of 34:34 to 10:10, and dataset 2 (source dataset) to dataset 1 (target dataset) in a ratio of 14:14 to 25:25. Due to the 1NN classifier's inability to predict classification probabilities, we do not use AUC (area under the curve) for performance measurement. Our approaches consistently outperform JDA and CORAL individually, regardless of the input features. We conduct experiments using five classic machine learning classifiers, applying the same sample separation. In the source dataset, we split the data into training and testing sets for parameter tuning, and then apply the trained model to the target dataset. The testing accuracy is reported along with the standard deviation in Table 2.

In the real-world case, the target labels are unknown, and therefore, the experiments presented in Table 3 were conducted using 20 random samples (instead of the previously mentioned 10:10 balanced approach) from the

Model	Dataset 1 → Dataset 2			Dataset 2 → Dataset 1		
	50 features	10 features	22 PCs	50 features	10 features	22 PCs
JDA	0.62 ± 0.11	0.72 ± 0.12	0.64 ± 0.06	0.57 ± 0.09	0.62 ± 0.06	0.61 ± 0.08
CORAL	0.60 ± 0.07	0.67 ± 0.06	0.61 ± 0.08	0.56 ± 0.04	0.56 ± 0.04	0.55 ± 0.06
GFK	0.55 ± 0.07	0.69 ± 0.10	0.58 ± 0.08	0.59 ± 0.04	0.61 ± 0.05	0.51 ± 0.03
EasyTL	0.66 ± 0.06	0.61 ± 0.08	0.50 ± 0.12	0.56 ± 0.09	0.56 ± 0.03	0.42 + 0.06
CEDA	**0.64 ± 0.09**	**0.76 ± 0.07**	**0.69 ± 0.07**	**0.59 ± 0.05**	**0.65 ± 0.04**	**0.62 ± 0.09**
SVM	0.49 ± 0.02	0.52 ± 0.05	0.52 ± 0.05	0.50 ± 0.01	0.55 ± 0.03	0.52 ± 0.04
LR	0.49 ± 0.06	0.56 ± 0.08	0.49 ± 0.06	0.50 ± 0.06	0.54 ± 0.02	0.50 ± 0.55
DT	0.57 ± 0.07	0.55 ± 0.05	0.50 ± 0.00	0.54 ± 0.04	0.48 ± 0.04	0.52 ± 0.02
KNN	0.48 ± 0.05	0.56 ± 0.04	0.42 ± 0.08	0.52 ± 0.03	0.58 ± 0.03	0.52 ± 0.03
RF	0.53 ± 0.04	0.55 ± 0.04	0.49 ± 0.06	0.56 ± 0.03	0.53 ± 0.02	0.52 ± 0.03
GBM	0.54 ± 0.02	0.61 ± 0.12	0.50 ± 0.05	0.50 ± 0.05	0.50 ± 0.04	0.50 ± 0.06
XGBoost	0.50 ± 0.03	0.57 ± 0.07	0.52 ± 0.05	0.49 ± 0.02	0.52 ± 0.05	0.50 ± 0.02

Table 2. Classification accuracy of two domain shifts on dataset 1 (171 samples) and dataset 2 (49 samples). Significant values are in bold.

Dataset 1 → Dataset 2	10 features	
	Accuracy	F1
JDA	0.73 ± 0.07	0.62 ± 0.12
CORAL	0.67 ± 0.07	0.59 ± 0.09
GFK	0.75 ± 0.10	0.62 ± 0.08
EasyTL	0.58 ± 0.11	0.52 ± 0.13
CEDA	**0.81 ± 0.07**	**0.67 ± 0.18**
SVM	0.39 ± 0.18	0.27 ± 0.24
LR	0.47 ± 0.16	0.33 ± 0.18
DT	0.47 ± 0.20	0.39 ± 0.26
KNN	0.54 ± 0.18	0.42 ± 0.21
RF	0.43 ± 0.15	0.48 ± 0.09
GBM	0.51 ± 0.05	0.43 ± 0.06
XGBoost	0.44 ± 0.10	0.36 ± 0.12

Table 3. Classification accuracy and F1 score using 10 filtered features. Significant values are in bold.

target samples as the testing datasets. The ratio of samples from the source dataset to the target dataset is 34:34 to 20. Additionally, we provide the F1 score to assess whether the model overfits the majority class.

Previously, we demonstrated how we selected 10 features and the feature score of each feature using mutual information. The provided feature scores indicate the contribution of each feature to the DA.

Conclusion and future work

This paper introduces a novel approach called CEDA for unsupervised domain adaptation (DA). CEDA is designed to align two domains by creating a domain-invariant feature representation. What sets our research apart from existing studies is that we address the challenges of small sample size and imbalanced healthcare data. Our model surpasses competing methods in accurately predicting fall risks for the target domain (new cohort) without relying on labeled data. In our future research, we plan to explore using signals directly instead of extracted features and incorporate a deep learning architecture to further enhance our approach.

Data availability

The data supporting this study's findings are available on request from Dr. Thurmon E. Lockhart, [thurmon.lockhart@asu.edu]. The data are not publicly available since data contains information that could compromise the privacy of research participants.

Received: 28 July 2023; Accepted: 8 February 2024
Published online: 12 February 2024

References

1. Cordts, M. *et al.* The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3213–3223 (2016).
2. Cortes, C. & Vapnik, V. Support-vector networks. *Mach. Learn.* **20**, 273–297 (1995).
3. Saxena, S. & Verbeek, J. Heterogeneous face recognition with CNNs. In *Computer Vision—ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8–10 and 15–16, 2016, Proceedings, Part III*. Vol. 14. 483–491 (Springer, 2016).
4. Klare, B. F., Bucak, S. S., Jain, A. K. & Akgul, T. Towards automated caricature recognition. In *2012 5th IAPR International Conference on Biometrics (ICB)*. 139–146 (IEEE, 2012).
5. Gong, B., Shi, Y., Sha, F. & Grauman, K. Geodesic flow kernel for unsupervised domain adaptation. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*. 2066–2073 (IEEE, 2012).
6. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **16**, 321–357 (2002).
7. Davar, N. F., de Campos, T., Windridge, D., Kittler, J. & Christmas, W. Domain adaptation in the context of sport video action recognition. In *Domain Adaptation Workshop, in Conjunction with NIPS* (2011).
8. Zhu, F. & Shao, L. Enhancing action recognition by cross-domain dictionary learning. In *BMVC* (2013).
9. Leggetter, C. J. & Woodland, P. C. Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models. *Comput. Speech Lang.* **9**, 171–185 (1995).
10. Reynolds, D. A., Quatieri, T. F. & Dunn, R. B. Speaker verification using adapted Gaussian mixture models. *Digit. Signal Process.* **10**, 19–41 (2000).
11. Chen, M., Xu, Z., Weinberger, K. & Sha, F. Marginalized denoising autoencoders for domain adaptation. arXiv Preprint arXiv:1206.4683 (2012).
12. Glorot, X., Bordes, A. & Bengio, Y. *Domain Adaptation for Large-Scale Sentiment Classification: A Deep Learning Approach*.
13. Wachinger, C., Reuter, M. & Initiative, A. D. N. Domain adaptation for Alzheimer's disease diagnostics. *Neuroimage* **139**, 470–479 (2016).
14. Cheplygina, V. *et al.* Transfer learning for multicenter classification of chronic obstructive pulmonary disease. *IEEE J. Biomed. Health Inform.* **22**, 1486–1496 (2017).
15. Csurka, G. Domain adaptation for visual applications: A comprehensive survey. arXiv Preprint arXiv:1702.05374 (2017).
16. Fernando, B., Habrard, A., Sebban, M. & Tuytelaars, T. Unsupervised visual domain adaptation using subspace alignment. In *Proceedings of the IEEE International Conference on Computer Vision*. 2960–2967 (2013).
17. Sun, B., Feng, J. & Saenko, K. Return of frustratingly easy domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 30 (2016).
18. Pan, S. J., Tsang, I. W., Kwok, J. T. & Yang, Q. Domain adaptation via transfer component analysis. *IEEE Trans. Neural Netw.* **22**, 199–210 (2010).
19. Long, M., Wang, J., Ding, G., Sun, J. & Yu, P. S. Transfer feature learning with joint distribution adaptation. In *Proceedings of the IEEE International Conference on Computer Visio.* 2200–2207 (2013).
20. Dai, W., Jin, O., Xue, G.-R., Yang, Q. & Yu, Y. *EigenTransfer: A Unified Framework for Transfer Learning.*
21. Xu, Y. *et al.* A unified framework for metric transfer learning. *IEEE Trans. Knowl. Data Eng.* **29**, 1158–1171 (2017).
22. Aljundi, R., Emonet, R., Muselet, D. & Sebban, M. Landmarks-based kernelized subspace alignment for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 56–63 (2015).
23. Wilson, G. & Cook, D. J. A survey of unsupervised deep domain adaptation. *ACM Trans. Intell. Syst. Technol. (TIST)* **11**, 1–46 (2020).
24. Long, M., Cao, Y., Wang, J. & Jordan, M. Learning transferable features with deep adaptation networks. In *International Conference on Machine Learning*. 97–105 (PMLR, 2015).
25. Sun, B. & Saenko, K. *Deep CORAL: Correlation Alignment for Deep Domain Adaptation*. Preprint http://arxiv.org/abs/1607.01719 (2016).
26. Kang, G., Jiang, L., Yang, Y. & Hauptmann, A. G. Contrastive adaptation network for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4893–4902 (2019).
27. Denton, E. L., Chintala, S. & Fergus, R. Deep generative image models using a Laplacian pyramid of adversarial networks. *Adv. Neural Inf. Process. Syst.* **28**, 133 (2015).
28. Kim, T., Cha, M., Kim, H., Lee, J. K. & Kim, J. Learning to discover cross-domain relations with generative adversarial networks. In *International Conference on Machine Learning*. 1857–1865 (PMLR, 2017).
29. Isola, P., Zhu, J.-Y., Zhou, T. & Efros, A. A. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1125–1134 (2017).
30. Maria Carlucci, F., Porzi, L., Caputo, B., Ricci, E. & Rota Bulo, S. Autodial: Automatic domain alignment layers. In *Proceedings of the IEEE International Conference on Computer Vision*. 5067–5075 (2017).
31. Li, Y., Wang, N., Shi, J., Hou, X. & Liu, J. Adaptive batch normalization for practical domain adaptation. *Pattern Recognit.* **80**, 109–117 (2018).
32. Karani, N., Chaitanya, K., Baumgartner, C. & Konukoglu, E. A lifelong learning approach to brain MR segmentation across scanners and protocols. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2018: 21st International Conference, Granada, Spain, September 16–20, 2018, Proceedings, Part I*. 476–484 (Springer, 2018).
33. AlBadawy, E. A., Saha, A. & Mazurowski, M. A. Deep learning for segmentation of brain tumors: Impact of cross-institutional training and testing. *Med. Phys.* **45**, 1150–1158 (2018).
34. Zhang, L. *et al.* Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation. *IEEE Trans. Med. Imaging* **39**, 2531–2540 (2020).
35. Lockhart, T. E., Smith, J. L. & Woldstad, J. C. Effects of aging on the biomechanics of slips and falls. *Hum. Factors* **47**(4), 708–729 (2005).
36. Doshi, K. B., Moon, S. H., Whitaker, M. D. & Lockhart, T. E. Assessment of gait and posture characteristics using a smartphone wearable system for persons with osteoporosis with and without falls. *Sci. Rep.* **13**(1), 538 (2023).
37. Lockhart, T. E. & Liu, J. Differentiating fall-prone and healthy adults using local dynamic stability. *Ergonomics* **51**(12), 1860–1872 (2008).
38. Boelens, C., Hekman, E. E. & Verkerke, G. J. Risk factors for falls of older citizens. *Technol. Health Care* **21**(5), 521–533 (2013).
39. Bizovska, L., Svoboda, Z., Janura, M., Bisi, M. C. & Vuillerme, N. Local dynamic stability during gait for predicting falls in elderly people: A one-year prospective study. *PloS one* **13**(5), e0197091 (2018).
40. Riva, F., Toebes, M. J. P., Pijnappels, M. A. G. M., Stagni, R. & Van Dieën, J. H. Estimating fall risk with inertial sensors using gait stability measures that do not require step detection. *Gait Posture* **38**(2), 170–174 (2013).
41. Koshmak, G. A., Linden, M., & Loutfi, A. Fall risk probability estimation based on supervised feature learning using public fall datasets. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 752–755 (IEEE, 2016).
42. Bargiotas, I. *et al.* Preventing falls: the use of machine learning for the prediction of future falls in individuals without history of fall. *J. Neurol.* **270**(2), 618–631 (2023).

43. Lockhart, T. E. *et al.* Prediction of fall risk among community-dwelling older adults using a wearable system. *Sci. Rep.* **11**(1), 20976 (2021).
44. Ferreira, R. N., Ribeiro, N. F. & Santos, C. P. Fall risk assessment using wearable sensors: A narrative review. *Sensors* **22**(3), 984 (2022).
45. Subramaniam, S., Faisal, A. I. & Deen, M. J. Wearable sensor systems for fall risk assessment: A review. *Front. Digit. Health* **4**, 921506 (2022).
46. Lockhart, T. E. *et al.* Prediction of fall risk among community-dwelling older adults using a wearable system. *Sci. Rep.* **11**, 20976 (2021).
47. Hotelling, H. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **24**, 417 (1933).
48. Kreer, J. A question of terminology. *IRE Trans. Inf. Theory* **3**, 208–208 (1957).
49. Walker, S. H. & Duncan, D. B. Estimation of the probability of an event as a function of several independent variables. *Biometrika* **54**, 167–179 (1967).
50. Quinlan, J. R. Induction of decision trees. *Mach. Learn.* **1**, 81–106 (1986).
51. Altman, N. S. An introduction to kernel and nearest-neighbor nonparametric regression. *Am. Stat.* **46**, 175–185 (1992).
52. Ho, T.K. Random decision forests. In *Proceedings of 3rd International Conference on Document Analysis and Recognition*. Vol. 1. 278–282 (IEEE Computer Society Press, 1995).
53. Friedman, J. H. Greedy function approximation: a gradient boosting machine. *Ann. Stat.* **3**, 1189–1232 (2001).
54. Chen, T., & Guestrin, C. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 785–794 (2016).
55. Wang, J., Chen, Y., Yu, H., Huang, M., & Yang, Q. Easy transfer learning by exploiting intra-domain structures. In *2019 IEEE International Conference on Multimedia and Expo (ICME)*. 1210–1215. (IEEE, 2019).

Author contributions

Z.G., T.W., and H.Y. conceived and designed the experiments, interpreted the results, and wrote the manuscript. T.E.L. and R.S. recruited participants, conducted the study, and contributed to writing and reviewing the manuscript. H.Y., Z.G., and T.W. conducted the machine-learning experiments and also contributed to the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to H.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.