This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2018.2839771, IEEE Journal of Biomedical and Health Informatics

Detecting Alzheimer's Disease on Small Dataset: A Knowledge Transfer Perspective

Wei Li. *Member, IEEE,* Yifei Zhao, Xi Chen, Yang Xiao, Yuanyuan Qin

***Abstract*—Computer-aided diagnosis (CAD) is an attractive topic in Alzheimer's disease (AD) research. Many algorithms are based on a relatively large training dataset. However, small hospitals are usually unable to collect sufficient training samples for robust classification. Although data sharing is expanding in scientific research, it is unclear whether a model based on one dataset is well suited for other data sources. Using a small dataset from a local hospital and a large shared dataset from the Alzheimer's Disease Neuroimaging Initiative (ADNI), we conducted a heterogeneity analysis and found that different functional magnetic resonance imaging (fMRI) data sources show different sample distributions in feature space. In addition, we proposed an effective knowledge transfer method to diminish the disparity among different datasets and improve the classification accuracy on datasets with insufficient training samples. The accuracy increased by approximately 20% compared with that of a model based only on the original small dataset. The results demonstrated that the proposed approach is a novel and effective method for CAD in hospitals with only small training datasets. It solved the challenge of limited sample size in detection of AD, which is a common issue but lack of adequate attention. Furthermore, the paper sheds new light on effective use of multi-source data for neurological disease diagnosis.**

***Index Terms*—Computer-aided diagnosis, Small dataset, Domain adaptation, Alzheimer's disease, rs-fMRI, Machine learning**

I. Introduction

THE problems associated with the aging population are becoming increasingly serious as people live longer and fertility rates decline in most countries. Furthermore, because a greater proportion of individuals are elderly, more people are at high risk of developing dementia. Currently, approximately 47 million people worldwide live with dementia, and this number is predicted to increase to more than 131 million by 2050 [1]. Alzheimer's disease (AD) is the most common form of dementia diagnosed in elderly people and significantly reduces their quality of life. An accurate and early diagnosis is essential for timely treatment and risk reduction. Over the past decade, several imaging modalities have been used in AD diagnosis, including diffusion tensor imaging (DTI) [2]–[6], structural magnetic resonance imaging (MRI) [7]–[12] and positron emission tomography (PET) [13]–[16]. Among these modalities, functional MRI (fMRI) plays an important role in monitoring brain activity and exploring the functional connectivity among different brain regions; therefore, fMRI is a promising methodology for the investigation and detection of brain disease [17], [18].

Researches about computer-aided diagnosis (CAD) systems which use machine-learning algorithms to diagnose or predict diseases have sprung up in recent years, especially for AD discrimination from fMRI scans [19]–[23]. However, certain problems remain in the development and application of CAD systems. Large datasets are highly important for AD research studies and CAD. Data sharing is a possible method for solving this problem. Some large research institutions and associations have begun to share their data, like the Alzheimer's Disease Neuroimaging Initiative (ADNI). However, even if large institutions are willing to share their data with smaller organizations, it is unclear whether the shared large dataset can be used to improve classification performance on the local small dataset. Many studies investigating AD discrimination using fMRI relied on a single data source with limited AD samples, including the above referenced researches. We found inconsistent or even contradictory conclusions in similar studies performing brain network analyses in AD. For example, Supekar et al. found that the global connectivity in patients with AD was higher than that in the normal controls (NCs), while Zhao et al. drew the opposite conclusion [24], [25]. In addition, Zhao et al. found that the average normalized characteristic path length in the patients with AD was greater than that in the NCs, while Sanz-Arigita et al. obtained the opposite result [25], [26]. We consider several factors that may have contributed to this phenomenon. First, different data sources may lead to dispersed sample distributions due to a mismatch in the age, race, and lifestyle of the subjects. Second, the machines and imaging parameters used for the data collection differed in the studies mentioned above. Third, the data preprocessing steps and related parameters differed. The impacts of the first and second factors are inevitable in research studies. Even in a single database, like ADNI, there are many institutions contributed to it with different imaging parameters (from 1.5T and 3.0T scanners for example). The research of Teipel et al. suggest that multi-site acquisition limits the use of rs-fMRI in AD diagnosis [27]. We assume that different fMRI data sources have different sample distributions, as illustrated in [28], and this may cause confusion. Thus, large shared dataset could not be directly mixed with a small dataset to obtain sufficient training samples for calibrating a classifier. In this situation, how can we improve the accuracy of classification using small

Wei Li, Yifei Zhao, Xi Chen, and Yang Xiao(corresponding author) are all with School of Automation, Huazhong University of Science and Technology, Wuhan, 430074, China, and also with Image Processing and Intelligent Control Key Laboratory of Education Ministry of China, Wuhan, 430074, P. R. China (E-mail:liwei0828@mail.hust.edu.cn; yifzh@hust.edu.cn; chenxi@hust.edu.cn; yang_xiao@hust.edu.cn).

Yuanyuan Qin is with Department of Radiology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, 430074, P. R. China.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2018.2839771, IEEE Journal of Biomedical and Health Informatics

datasets in local hospitals?

Transfer learning has become a "hot topic" for solving classification problems using multi-source data in the research field of computer vision [29]–[31]. Some researchers are aware of this development and have been exploring the application of this method to biomedical analysis. Van Opbroek et al. used transfer learning in supervised image segmentation of the brain using multi-site data and obtained good results [32]. Conjeti et al. proposed a supervised domain adaptation (a sort of transfer learning conditions) framework that combined domain alignment, feature normalization, and leaf posterior reweighting to realize the in vivo characterization of atherosclerotic tissues in the presence of a distribution shift between the training and testing data and demonstrated that their approach is a superior alternative [33]. Wachinger and Reuter (integrated instance weighting into an elastic-net multinomial regression to mitigate overfitting and correct variations among large multi-center datasets and showed that when some samples from the target dataset were included in the training set, the classification accuracy was improved compared with that when the training solely relied on the target dataset [34]. Goetz et al. (2016) combined reweighting observations and random forests to correct sampling selection errors introduced by sparse annotations and reduced the labeling time by a factor of more than 70 without sacrificing accuracy [35]. As demonstrated by the above-mentioned studies, domain adaptation is a promising approach in medical signal processing. Although great progress has been achieved in transfer learning, to the best of our knowledge, no studies investigating a target dataset with a very limited number of samples. Aforementioned researches are all with relatively large sample size in both source dataset and target dataset. Once the dataset size is too small, it will renders these developed methods unsuitable for solving the problem of classifier calibration and may greatly affect the practical application of transfer learning techniques. Thus, it's highly significant to effectively use large datasets from different data sources for classifier calibration on limited sample size, particularly for CAD systems in small hospitals and organizations.

In this study, we first explored the heterogeneity of AD datasets from different data sources and confirmed our hypothesis that different fMRI data sources have different sample distributions in the feature space. Then, a simple and efficient domain adaptation method was used to diminish the disparity in the sample distribution between the large shared dataset and the small current dataset. Finally, a common machine-learning model was built for AD discrimination based on the data after adaptation. Using this procedure, we achieved a significant increase in the accuracy of classification using only the small datasets. More specifically, the classification accuracy was more than 20% greater than that using classifiers trained with naïve combinations of samples from different data sources and approximately 30% greater than that using classifiers trained only with the small target dataset. Wachinger and Reuter used general linear model integrated with instance weighting in domain adaptation of AD classification and obtained improved performance [34]. The same method has been applied in this study for comparison. However, we did not find improvement in the classification accuracy compared with that of using naïve combinations of samples. It indicates that the performance of aforementioned studies will be degraded more or less in such a condition of very limited training samples. The results demonstrated that the solution proposed in this paper is an effective method for overcoming the challenges of CAD using small sample sets and paves a new way for individual hospitals and organizations to build specific auxiliary diagnosis applications.

II. Methods

We introduced a simple novel domain adaptation method designed to diminish the disparity in the sample distributions among datasets from different sources. The framework of our method is described in Fig. 1. First, we extracted weighted connections between different functional regions as original features through brain network modeling. Second, a feature selection step was executed before the domain adaptation because the dimensions of the original features were too large. The selected features from the two different data sources were distributed in two separate feature spaces. Then, we performed a modified subspace alignment to align the sample points from the two separate feature spaces into the same subspace. Finally, the aligned samples in one subspace were as an integrated dataset for the classifier training and testing.

A. Modified Subspace Alignment Method

The domain adaptation method proposed in this paper was adapted from Fernando et al. [36]. They proposed a subspace alignment method for visual recognition that yielded good results in an object recognition task, and this algorithm is efficient due to its intrinsic simplicity. Here, we modified this method to make it suitable for a small dataset and fMRI data.

Given two datasets from different domains, i.e., source domain D_S containing m samples and target domain D_T containing n samples, $[x_1, \ldots, x_m] = X_S \in D_S$ and $[x_1, \ldots, x_n] = X_T \in D_T$. The subspace alignment mapped the data into subspaces and learned transfer matrix M, which aligned one subspace with the other. Assuming that the feature dimension of the samples is d_m, first, we can obtain d eigenvectors for each domain by composing their subspaces $V_S^{d_m \times d}$ and $V_T^{d_m \times d}$ using singular value decomposition (SVD). then each data point was projected into its respective subspace by multiplying $V_S^{d_m \times d}$ or $V_T^{d_m \times d}$ as appropriate. The function of mapping the feature vectors from one domain to the subspace of the other domain is learned by minimizing the following Bregman matrix divergence:

$$F(M) = \| V_S M - V_T \|_F^2 \tag{1}$$

$$M^* = arg\, min_M(F(M)) \tag{2}$$

where $\|\cdot\|_F^2$ is the Frobenius norm. To obtain the optimal M^*, equation 1 can be rewritten as follows:

$$F(M) = \| V_S' V_S M - V_S' V_T \|_F^2 = \| M - V_S' V_T \|_F^2 \tag{3}$$

In this equation, we can obtain the optimal M^* as $M^* = V_S' V_T$.

Once the feature vectors are mapped, all the data are aligned into a single subspace. Finally, all projected feature vectors belonging to a certain subspace can be used for the training and prediction. The entire procedure is presented in Algorithm 1.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2018.2839771, IEEE Journal of Biomedical and Health Informatics

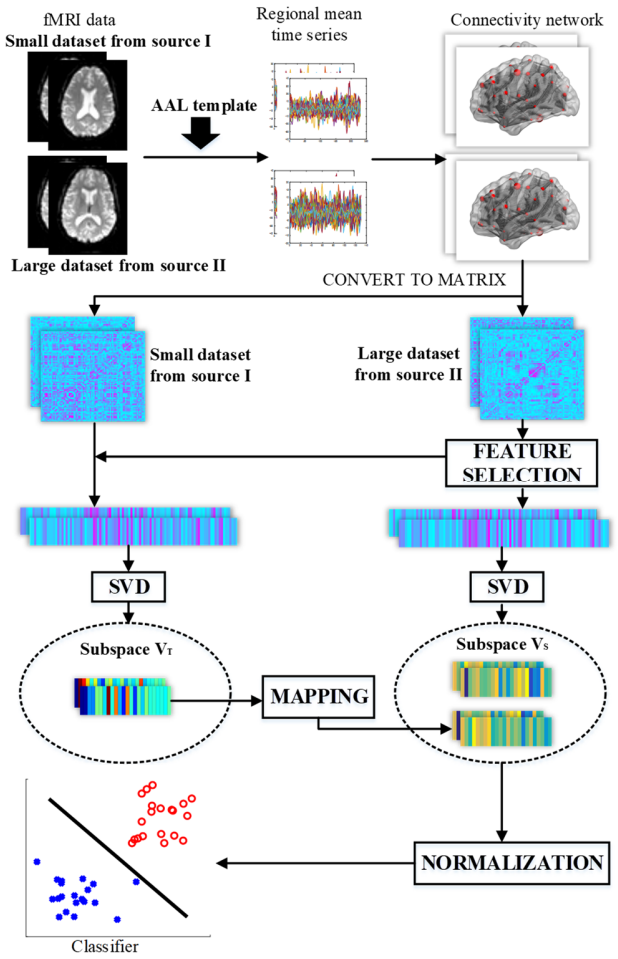


Fig. 1. Framework of the classification using the modified subspace alignment method.

Using the subspace alignment method, the performance is related to the sole hyperparameter d. Parameter d corresponds to the dimension of the subspaces that has the limitation of being smaller than the number of samples or the feature dimension d_m. According to consistency theorem on similarity function of source data and target data, we can use theoretical result to determine the upper bound of d. Given a fixed deviation $\gamma > 0$, with the inequation $\left\| V_S^d M V_T^{d'} - V_{S_n}^d M_n V_{T_n}^{d'} \right\| \leqslant \gamma$, a subset of $d \in \{d|1, \ldots, d_{max}\}$ can be obtained. That means as long as we select a subspace dimension d satisfying the condition $d \leq d_{max}$, the solution mapping procedure is stable and not over-fitting. In this study, since the range of d is not wide, we have tried all the possible value of d and got the optimal one.

Compared with the method in [36], we replaced the principal component analysis (PCA) with SVD, which has several advantages. First, the value of the features used in this situation ranged from zero to one, and many features in the feature vectors had a value close to zero. Under this circumstance, calculating the covariance matrix in a PCA may lose numbers that are very close to zero, while using SVD can obtain more stable results. Second, the dimension of the subspace cannot be larger than the number of samples in the smaller dataset or feature dimension d_m if using a PCA. In fact, the number of samples is very limited and far smaller than d_m in our situation, but SVD is free of the constraint of insufficient samples, and we used right singular vectors as eigenvectors. Finally, performing a PCA requires centered and standardized data, which is not necessary in SVD.

Algorithm 1. Subspace alignment using SVD
1. Performing SVD on X_S and X_T via $X_S = U_S \Sigma_S V_S'$ and $X_T = U_T \Sigma_T V_T'$, respectively.
2. Determining dimension d of the subspace by keeping the first d columns of V_S and V_T yields the mapping matrices $V_S^{d_m \times d}$ and $V_T^{d_m \times d}$.
3. Mapping the feature vectors of all samples to the subspace of the source domain as follows: $S_S^{m \times d} = X_S^{m \times d_m} V_S^{d_m \times d}$ and $T_S^{n \times d} = X_T^{n \times d_m} V_T^{d_m \times d} V_T'^{d \times d_m} V_S^{d_m \times d}$.
4. Training a classifier with $S_S^{m \times d}$ and then using this classifier to disambiguate $T_S^{n \times d}$.

B. *Discriminant Analysis Classifier*

A discriminant analysis classifier is a nonparametric probabilistic model used to classify a new observation based on the following three quantities: posterior probability, prior probability, and cost. This model assumes that the data in each class have a multivariate normal distribution. The model uses the same covariance matrix for the liner discriminant analysis and individual covariance of each class for the quadratic analysis. This approach has been successfully applied for the detection of AD in [37].

Assuming there are K classes and given observation x, the predicted classification is the minimization of the expected classification cost as follows:

$$\hat{y} = \arg\min_{y=1,\ldots,K} \sum_{k=1}^{K} \hat{P}(k|x) C(y|k) \quad (4)$$

where $\hat{y}$ is the predicted classification, $\hat{P}(k|x)$ is the posterior probability of class K for observation x, and $C(y|k)$ is the cost of classifying an observation as y when its true class is k. Generally, the cost is 0 if the classification is accurate and 1 otherwise. A thorough presentation of discriminant analysis can be found in [38]. We have tried both type of discriminant in the study, and better performance achieved with linear discriminant. In this paper, all the experiments and results were based on linear discriminant.

III. Experimental Setup

A. *Data collection and preprocessing*

There were two datasets used in this study. Data source I, the local small dataset (target dataset), was collected in Tongji Hospital at Wuhan. It contained 26 subjects in total, including 12 patients diagnosed with AD according to the NINCDS -ADRDA criteria and 14 healthy controls [39]. Each subject had one fMRI scan only. These 26 subjects are all available samples in Tongji Hospital. Data source II, the large shared dataset (source dataset), contained 86 subjects. Data source II used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative database

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2018.2839771, IEEE Journal of Biomedical and Health Informatics

TABLE I
DEMOGRAPHIC INFORMATION AND OTHER SAMPLE CHARACTERISTICS

Data sources	Group	AD	NC
ADNI	Female/male	56/61	98/77
	Age	74.6±7.5	75.5±6.1
	MMSE	21.3±3.5	28.9±1.5
Tongji	Female/male	6/6	8/6
	Age	65.7±11.9	65.7±7.5
	MMSE	16.7±3.0	28.5±1.2

(adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early AD. For up-to-date information, see www.adni-info.org. Most subjects from data source II had undergone fMRI scan more than once; there were 117 scans of 34 AD patients and 175 scans of 52 healthy controls (excluding scans with abnormal head motion and different data acquisition setups) in total. Each scan was regarded as an individual sample to make the shared dataset as large as possible. To be noted, the scans which belong to one subject were regarded as a whole and either all distributed to training set or testing set. It won't happen where some scans from a subject were in training data while some other scans from the same subject were in testing set. Therefore, it's a feasible way to enlarge the training set.

The study protocol was approved by the ethics committee at Tongji medical college of Huazhong University of Science and Technology, and written informed consents were obtained from all subjects from data source I. Demographic information and other characteristics of the two datasets are shown in Table I.

We collected the Tongji dataset on a GE signa HDxt 3.0 Tesla MRI scanner. The specific parameters are as follows: TR/TE = 2000/30 ms, flip angle = 90°, imaging matrix = 64×64, voxel size = 3.0 mm × 3.0 mm × 3.0 mm, number of slices =33, and each series has 240 volumes. fMRI data in the ADNI database was acquired on a 3-T Philips MRI scanner with the following parameters: TR/TE = 3000/30 ms, flip angle = 80°, imaging matrix = 64 × 64, voxel size = 3.31 mm × 3.31 mm × 3.31 mm, number of slices =48, and each series has 140 volumes.

Post-processing of all fMRI images was performed using SPM8 (http://www.fil.ion.ucl.ac.uk/spm), REST and DPARSF [40], [41]. After removing the first ten volumes of each series for signal equilibration, slice timing and realigning for head motion correction were carried out. Then the functional images were normalized into the Montreal Neurological Institute (MNI) space using echo-planar imaging (EPI) template. Samples with head motion larger than 2.5mm were excluded. The resulting images were spatially smoothed using a Gaussian kernel with 6 mm × 6 mm × 6 mm FWHM, and REST was used to remove the linear trends of time courses. Finally, temporal band-pass filtering within the interval of 0.01–0.08 Hz was applied to the time courses of each voxel, and nuisance covariates, including six head motion parameters, global mean signal, white matter signal, and cerebrospinal fluid sign were regressed out. Time series of 90 regions of interests (ROIs) for each sample were extracted by warping the automated anatomical labeling (AAL) atlas for processed brain images for further analysis [42].

B. Feature extraction

We calculated the pairwise Pearson correlation coefficients between any two time-series of 90 ROIs in each scan, resulting in a 4005-dimensional feature vector for one scan describing the correlation coefficient matrix. Then the Kendall's tau correlation coefficients were used for feature selection procedure [43]. The Kendall's tau coefficients has been proved to be effective for feature selection in neuroimaging studies about CAD [19], [44]. Since the number of samples is limited (292 from the ADNI dataset and 26 from the Tongji dataset), we conducted the experiments with 10 to 200 selected features at an interval of 10 to evaluate the generalizability of the method. In the experiments of adaptation, we first performed the feature selection procedure in large ADNI dataset. Then, the consequential selection was applied to the small Tongji dataset. That means we obtained a subset of features in ADNI dataset after feature selection procedure and the same subset of features was chosen for Tongji dataset.

C. Experimental procedure

To evaluate the effectiveness of our method, the following sets of AD discrimination experiments were performed:

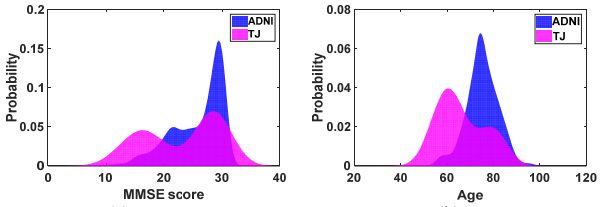
(i). Only Tongji: Classification task only with local Tongji dataset. During this task, leave-one-out cross-validation (LOOCV) strategy was applied with 26 samples of Tongji dataset. More specifically, each time 25 samples were used for feature selection and training the discriminant analysis classifier, and the remaining one was used for testing;

(ii). Naïve Combination: ADNI dataset was added to training set in Experiment (i). More specifically, each time 25 samples from Tongji dataset plus all the ADNI samples were used for feature selection and training the discriminant analysis classifier, and the remaining one in Tongji dataset was used for testing;

(iii). With Adaptation: As illustrated in Fig. 3. At first, feature selection was performed in ADNI dataset and a subset of features was obtained in this procedure, then the same subset of features was chosen for each sample for Tongji dataset. After that, Tongji data (target domain) was projected to its subspace and mapped to the subspace of ADNI data (source domain). ADNI data was projected to its own subspace. With this, ADNI data and Tongji data were aligned in the same subspace. Finally, the ADNI and 25 out of 26 Tongji samples in ADNI subspace were used for training the discriminant analysis classifier, and the remaining one for testing. The final step is similar to Experiment (ii), where the difference is that the data in (iii) was after adaptation.

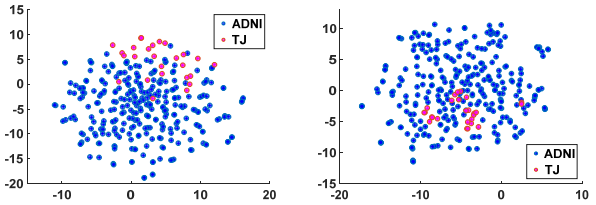
(iv). Instance Weighting: In addition, a general linear model integrated with instance weighting for AD classification was built for comparison, which is the method performed in [34]. Tongji dataset was split into two part, training part and testing part. For each sample in training part, a weight was evaluated according to probability of these samples in ADNI domain. Then these samples along with ADNI samples were fed into a general linear model with log-likelihood function and elastic-net multinomial regression. All the fed samples in log-likelihood function were weighted. The weights of the training part from Tongji dataset were as evaluated at beginning,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2018.2839771, IEEE Journal of Biomedical and Health Informatics



(a) MMSE (b) Age

Fig. 2. The probability distributions of the MMSE scores and ages in the two datasets. (a) The probability distributions of the MMSE scores. (b) The probability distributions of ages.



(a) Before adaptation (b) After adaptation

Fig. 3. Visualization of the distribution of the sample points from the two separate data sources. (a) The distribution before adaptation. (b) The distribution after adaptation with modified subspace alignment.

and the ADNI samples got a constant weight equal to 1. Finally, the trained general linear model was used to classify the testing part of Tongji dataset. The splitting procedure was repeated for 50 times as we randomly sampled the training set each time. Due to the limited number of samples in Tongji domain, we added fifty percent of the Tongji samples to the training set instead of ten to thirty percent in [34].

The accuracy, sensitivity, specificity and area under the curve (AUC) in each experiment were calculated. All experiments were performed with Matlab R2016a.

IV. Results

The mixture of two datasets are heterogeneous. Fig. 2 illustrates the probability distributions of mini-mental state examination (MMSE) scores and ages in both datasets. These distributions were estimated using a kernel density estimation. In addition, a 4005-dimension feature per scan was extracted, and the t-SNE method (developed from Stochastic Neighbor Embedding) was used to decrease the feature dimension and enable the visualization of high dimensional data [45]. The result is shown in Fig. 3(a). Obviously, the sample points from the Tongji dataset cluster mainly in the upper margin of the ADNI sample points before adaptation. It indicates that the different fMRI data sources show different sample distributions in the feature space. Consistency in feature distribution of samples is of crucial importance for achieving good classification performance in pattern recognition. Therefore, it is necessary for CAD that uses fMRI data from multiple sources to eliminate the inconsistency in the sample distribution across different data sources. The sample distribution of the two datasets was also visualized in a 2-dimension feature space after adapting the Tongji data to the domain of ADNI dataset. As shown in Fig. 3(b), sample points from the Tongji dataset no longer lay in the margins of the ADNI data points but shifted to a distribution consistent with that of ADNI dataset. The modified subspace alignment method bridged the gap of domain differences to a certain extent.

The classification task only with the local small dataset from Tongji data source did not perform well, as shown in Fig. 4. Its accuracy was slightly greater than 50%, which is almost like random guesses. It's difficult to construct a robust model for auxiliary diagnosis of AD in a sample set as small as Tongji dataset. When we used the large shared dataset from ADNI and mixed with Tongji dataset, the performance of the classifier significantly increased, reaching an accuracy of 60%. Thus, enlarging the size of training set indeed improved the performance of a classifier. However, this improvement was inadequate compared with the classification results obtained in the adaptation experiment, in which the accuracy was greater than 80%. The unsatisfactory accuracy before adaptation indicates that it is not possible to obtain reliable classification results by simply adding different data sources to obtain a larger training set.

The specific performance indicators are summarized in Table II. The accuracy after the adaptation increased by nearly thirty percent compared with that in the classification experiment using only Tongji dataset. However, the result of the general linear model integrated with instance weighting is not promising, for its performance was not improved compared with that of naïve combining the two datasets to train the classifier. Thus, ordinary domain adaptation methods may not be well suited for the situation in which the target samples are very insufficient. More considerations should be taken on solving this challenge.

The number of selected features influences the performance of the AD classification. As shown in Fig. 4, when we selected a few features for the classification, the classifier performed poorly. The accuracy of the classification improved as the number of selected features increased. However, the classifier performance stopped improving once a certain number of features (depending on the sample size) was reached. Too few features are insufficient for classification, while too many features may be redundant. This finding is consistent with previous studies [46]. We selected the features by ranking the Kendall's tau correlation coefficients. Table III lists the

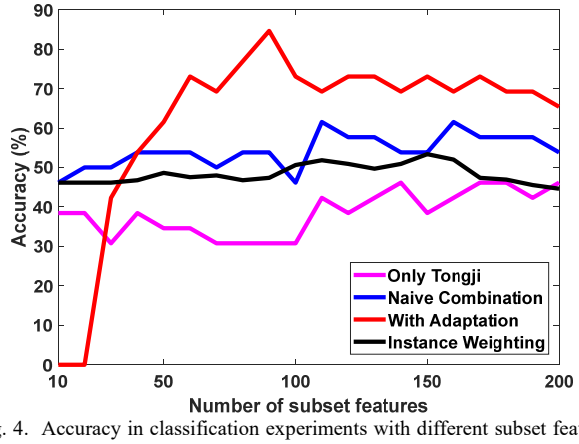


Fig. 4. Accuracy in classification experiments with different subset features when d = 25. Black line: accuracy for general linear model integrated with instance weighting. The missing accuracy in beginning of red line is due to the limitation that the feature dimension should be great than subspace dimension d.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2018.2839771, IEEE Journal of Biomedical and Health Informatics

TABLE II
BEST CLASSIFICATION RESULTS OF ALL EXPERIMENTS*

Experiment	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC (%)
Only Tongji	49.0	42	57	32
Naive combination	61.5	58	64	63
Instance weighting	55.3±2.9	62±2.5	55±2.7	60±3.0
With adaptation	**84.6**	**92**	**79**	**80**

* The best results were obtained when the subspace dimension d = 25. The +/- STD in Instance weighting was for the reason the experiment was repeated for 50 times as we randomly sampled the training set each time. In other classification experiments the LOOCV strategy was applied.

TABLE III
FEATURES WITH |τ| SCORE RANKING IN THE TOP TEN

Feature(connection between regions)	|τ|
Hippocampus_R↔Frontal_Sup_L	0.254
Hippocampus_R↔Frontal_Mid_L	0.249
Temporal_Inf_R↔Heschl_R	0.238
Occipital_Mid_L↔Precentral_L	0.231
Cuneus_L↔Cingulum_Mid_R	0.230
Postcentral_R↔Frontal_Sup_L	0.228
Temporal_Inf_L↔Occipital_Inf_L	0.226
Postcentral_R↔Frontal_Mid_L	0.223
Temporal_Inf_L↔Fusiform_L	0.223
Paracentral_Lobule_L↔Occipital_Mid_L	0.219

features with a |τ| score ranking in the top ten in Experiment (iii). The top 50 features selected from the ADNI dataset are labeled in Fig. 5, which was generated with Brain-Net Viewer [47].

Subspace dimension d plays an important role in the domain adaptation. The accuracy increased as subspace dimension d increased as shown in Fig. 6. We have compared the SVD subspace alignment with original method in all values of d with 90 selected features. The performance of SVD method is as good as that of original one in most cases, and sometimes it's even better. The optimal d is 25 in our experiment. Due to the limitation of sample size, we have tried all the possible values. In other applications, the parameter d can be tuned according to consistency theorem described in modified subspace alignment method.

V. DISCUSSION

CAD systems used for AD classification based on fMRI data have become a major research topic in recent years. However, it is difficult for small hospitals and research organizations to collect sufficient samples for accurate classification. Many small hospitals and research organizations frequently encounter this difficulty. Data sharing is a possible choice for solving this problem. In this study, because the fMRI dataset from Tongji Hospital was too small, the classifier trained with limited samples performed poorly.

The ADNI dataset was introduced to enlarge the training dataset and improve the classification accuracy. Most studies investigating AD identification conducted their experiments using a single dataset [19]–[21], [23], [37], [46], [48]–[51]; thus, their models usually performed well for samples selected from the same specific data source. However, it is unclear whether a model built on a particular dataset is well suited for different data sources. The model trained with the ADNI dataset did not perform well on the Tongji dataset. Although a naïve augmentation with ADNI data improved the classification accuracy, the classification performance remained poor for further practical application. We think the problem laid in the difference between the two data sources where the samples were obtained. By performing a heterogeneity analysis of the different data sources, we found that the feature distribution of the samples from Tongji Medical College differed from that of the ADNI data, which confirmed our assumption. This discrepancy may relate to many factors. Obviously, the probability distribution of age and MMSE score in Tongji dataset is different from that in ADNI dataset, as shown in Fig. 2 and Table I. Furthermore, the equipment and acquisition parameters are different between the two datasets. These factors result in heterogeneity in data from different sources and the failure of naïve combination of multi-site data.

Regarding multi-source data utilization, transfer learning is a potential way in medical signal analysis. Heimann et al. applied instance weighting to the image fusion of trans-esophageal echography (TEE) and X-ray fluoroscopy for ultrasound transducer localization [52]. Schlegl et al. used data from multiple sites via a semi-supervised learning approach to

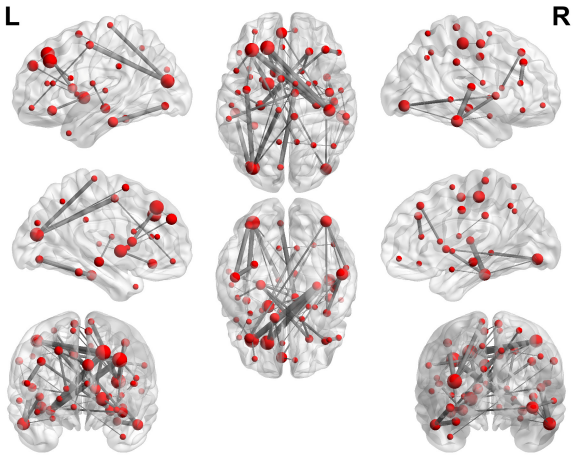


Fig. 5. Top 50 selected features from the ADNI dataset in Experiment (iii). The size of the node is proportional to the number of significant connections with which the node (ROI) is involved. The thickness of the edge is proportional to the value of the |τ| score of the features.

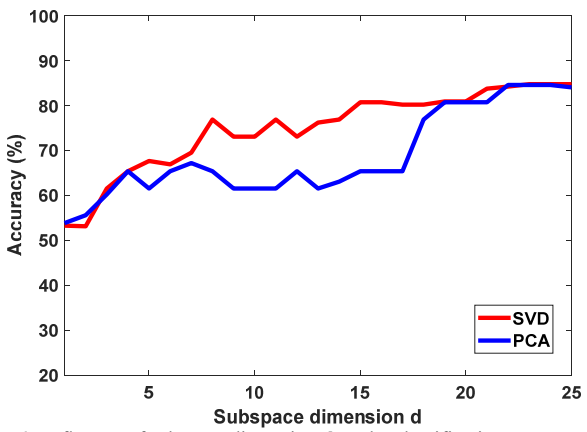


Fig. 6. Influence of subspace dimension d on the classification accuracy when the number of subset features was 90. The red line denotes the accuracy of the classification with SVD (modified subspace alignment method), and the blue line denotes the accuracy of the classification with PCA (original subspace alignment method).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2018.2839771, IEEE Journal of Biomedical and Health Informatics

improve lung tissue classification [53]. Van Opbroek et al. showed that transfer learning outperformed common supervised learning approaches in MRI brain-segmentation tasks using multi-site data [32]. Domain adaptation is a promising approach in transfer learning. In this study, we used a simple domain adaptation method to diminish the disparity in the feature distribution between the large and small datasets for AD detection. Our experiments demonstrated that adapting the Tongji dataset to the ADNI domain using this method yielded a dramatic improvement in the classification. The accuracy was nearly 20% greater than that using the naïve combination of samples and approximately 30% greater than that using the small dataset. Our domain adaptation method diminishes the disparity in the feature space between samples from different data sources and effectively improves the classification accuracy when sample size is extremity limited. This study provides a novel method for small hospitals and research organizations to develop CAD systems for neurological diseases.

The modified subspace alignment method introduced in this paper is a relatively easy algorithm compared with other domain adaptation methods, such as feature augmentation based approaches, feature transformation based approaches, dictionary based approaches, and other methods [54]–[58]. This method can be conveniently adapted for multiple domains as long as we map the other domains to a single source domain. Information loss is a limitation of the subspace alignment method because this method employs SVD to produce the base vectors of the subspaces. Only hyperparameter d slightly influenced the experimental results. Hyperparameter d denotes how many dimensions there are in a subspace, and d is directly concerned with information loss. Hence, the chart of d is in accordance with our expectations. The subspace alignment method has been successfully applied in other medical image recognition applications, demonstrating its simple form and extensive application [33].

The general liner model integrated with instance weighting yielded disappointing results. Its overall performance was not better than that of the naïve combination method. We think the insufficient samples in target dataset is the major reason. The instance weighing estimates the probability distributions in the source dataset and target dataset and then evaluates the probability of the target samples under the source distribution. The limited number of samples in the target dataset complicates the precise estimation of the target distribution; therefore, it brought large error when evaluating the weights. However, the modified subspace alignment method uses all data in the target dataset regardless of whether their labels are available and adapts the target samples to the subspace of the source domain, thereby decreasing the disparity between the two datasets. Our method is well suited for classification task on a very small dataset.

In this study, the feature selection strategy is based on Kendall's tau correlation coefficients. The top ten selected features and the regions involved are listed in Table IV. Among these regions, the hippocampus is ranked first and has been demonstrated to be highly related to AD [18], [59], [60]. The connectivity between frontal areas and other regions, which was chosen as a classification feature, has been associated with AD in other studies [19], [61]. Additionally, most other listed regions are consistent with several previous studies [19], [60], [62]–[65]. The consistency of our results with those obtained in other studies suggests that the features selected with Kendall's tau correlation coefficients are reliable for classification.We only considered functional connectivity features because these features are widely used in studies investigating AD identification based on fMRI [19], [37], [46], [66]. Other types of biomarkers or features can be included in this method to achieve better classification performance.

There are some limitations to our study. We only demonstrated that different fMRI data sources show different distributions in the feature space. However, how factors, such as devices, the setup of the technical parameters, etc., influence the sample distribution in the feature space remains unclear. On the other hand, AD studies progress slowly in local hospitals, and it is difficult for us to obtain additional samples for analysis and comparison. Meanwhile, physicians in these developing regions do not have as many rich experiences as those doctors in large medical centers, thereby increasing the need of CAD assistance. For CAD systems on AD classification, more studies on early AD detection are now trending. MCI is an important early stage of Alzheimer's disease, and distinguishing MCI from normal aging is more significant and difficult than detecting AD. There are many excellent papers about MCI classification, and we have got a lot inspiration from these studies [67]–[69]. To identity MCI population is one of the important task in our future works. As the develop of deep learning, using deep neural network in auxiliary diagnosis to automatic extract features and conduct transfer learning is promising. We are looking forward to building robust deep learning architecture in the future.

VI. Conclusion

In this paper, we demonstrated that the AD classification task using a small dataset can be better solved using the modified subspace alignment method. This method can effectively improve the accuracy of the classification in small sample sets. Researchers can use this method to relieve the challenge of extremely limited sample size, particularly when collecting neuroimaging data is difficult and computer-aided diagnoses with limited samples are required. Our work may also assist researchers to make better use of shared data and promote the exchange of collected data.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (61473131, 61502187, 81401389), International Science & Technology Cooperation Program of Hubei Province, China (Grant No. 2017AHB051), and HUST Interdisciplinary Innovation Team Foundation (Grant No. 2016JCTD120).

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's

Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

REFERENCES

[1] M. Prince, A. Wimo, M. Guerchet, A. Gemma-Claire, Y.-T. Wu, and M. Prina, "World Alzheimer Report 2015: The Global Impact of Dementia - An analysis of prevalence, incidence, cost and trends," *Alzheimer's Dis. Int.*, p. 84, 2015.

[2] M. Graña *et al.*, "Computer Aided Diagnosis system for Alzheimer Disease using brain Diffusion Tensor Imaging features selected by Pearson's correlation," *Neurosci. Lett.*, vol. 502, no. 3, pp. 225–229, 2011.

[3] M. Dyrba *et al.*, "Combining DTI and MRI for the automated detection of Alzheimer's disease using a large European multicenter dataset," in *International Workshop on Multimodal Brain Image Analysis*, 2012, pp. 18–28.

[4] S. Haller *et al.*, "Individual classification of mild cognitive impairment subtypes by support vector machine analysis of white matter DTI," *Am. J. Neuroradiol.*, vol. 34, no. 2, pp. 283–291, 2013.

[5] W. Lee, B. Park, and K. Han, "Classification of diffusion tensor images for the early detection of Alzheimer's disease," *Comput. Biol. Med.*, vol. 43, no. 10, pp. 1313–1320, 2013.

[6] T. M. Nir *et al.*, "Diffusion weighted imaging-based maximum density path analysis and classification of Alzheimer's disease," *Neurobiol. Aging*, vol. 36, no. S1, pp. S132–S140, 2015.

[7] R. Cuingnet *et al.*, "Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database.," *Neuroimage*, vol. 56, no. 2, pp. 766–81, 2010.

[8] E. Westman, J.-S. Muehlboeck, and A. Simmons, "Combining MRI and CSF measures for classification of Alzheimer's disease and prediction of mild cognitive impairment conversion," *Neuroimage*, vol. 62, no. 1, pp. 229–238, 2012.

[9] C. Aguilar *et al.*, "Different multivariate techniques for automated classification of MRI data in Alzheimer's disease and mild cognitive impairment," *Psychiatry Res. - Neuroimaging*, vol. 212, no. 2, pp. 89–98, 2013.

[10] A. Ortiz, J. M. Górriz, J. Ramírez, F. J. Martínez-Murcia, and A. D. N. Initiative, "LVQ-SVM based CAD tool applied to structural MRI for the diagnosis of the Alzheimer's disease," *Pattern Recognit. Lett.*, vol. 34, no. 14, pp. 1725–1733, 2013.

[11] L. Khedher, J. Ramírez, J. M. Górriz, A. Brahim, and F. Segovia, "Early diagnosis of Alzheimer's disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images," *Neurocomputing*, vol. 151, no. P1, pp. 139–150, 2015.

[12] F. de Vos *et al.*, "Combining multiple anatomical MRI measures improves Alzheimer's disease classification," *Hum. Brain Mapp.*, vol. 37, no. 5, pp. 1920–1929, 2016.

[13] N. L. Foster *et al.*, "FDG-PET improves accuracy in distinguishing frontotemporal dementia and Alzheimer's disease," *Brain*, vol. 130, no. 10, pp. 2616–2635, 2007.

[14] K. Herholz, S. F. Carter, and M. Jones, "Positron emission tomography imaging in dementia," *Br J Radiol*, vol. 80, no. Special_Issue_2, pp. S160-167, 2007.

[15] D. Zhang, Y. Wang, L. Zhou, H. Yuan, and D. Shen, "Multimodal classification of Alzheimer's disease and mild cognitive impairment," *Neuroimage*, vol. 55, no. 3, pp. 856–867, 2011.

[16] D. Zhang and D. Shen, "Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease," *Neuroimage*, vol. 59, no. 2, pp. 895–907, 2012.

[17] E. L. Dennis and P. M. Thompson, "Functional brain connectivity using fMRI in aging and Alzheimer's disease," *Neuropsychology Review*, vol. 24, no. 1. pp. 49–62, 2014.

[18] J. L. O'Brien *et al.*, "Longitudinal fMRI in elderly reveals loss of hippocampal activation with clinical decline," *Neurology*, vol. 74, no. 24, pp. 1969–1976, 2010.

[19] E. Challis, P. Hurley, L. Serra, M. Bozzali, S. Oliver, and M. Cercignani, "Gaussian process classification of Alzheimer's disease and mild cognitive impairment from resting-state fMRI," *Neuroimage*, vol. 112, pp. 232–243, 2015.

[20] X. Chen, H. Zhang, Y. Gao, C. Y. Wee, G. Li, and D. Shen, "High-order resting-state functional connectivity network for MCI classification," *Hum. Brain Mapp.*, vol. 37, no. 9, pp. 3282–3296, 2016.

[21] A. Khazaee, A. Ebrahimzadeh, and A. Babajani-Feremi, "Classification of patients with MCI and AD from healthy controls using directed graph measures of resting-state fMRI," *Behav. Brain Res.*, 2016.

[22] S. H. Hojjati, A. Ebrahimzadeh, A. Khazaee, and A. Babajani-Feremi, "Predicting conversion from MCI to AD using resting-state fMRI, graph theoretical approach and SVM," *J. Neurosci. Methods*, vol. 282, pp. 69–80, 2017.

[23] R. Armananzas, M. Iglesias, D. A. Morales, and L. Alonso-Nanclares, "Voxel-based diagnosis of Alzheimer's disease using classifier ensembles," *IEEE J Biomed Heal. Inf.*, vol. XX, no. X, pp. 1–7, 2016.

[24] K. Supekar, V. Menon, D. Rubin, M. Musen, and M. D. Greicius, "Network analysis of intrinsic functional brain connectivity in Alzheimer's disease," *PLoS Comput. Biol.*, vol. 4, no. 6, 2008.

[25] X. Zhao *et al.*, "Disrupted small-world brain networks in moderate Alzheimer's disease: A resting-state fMRI study," *PLoS One*, vol. 7, no. 3, 2012.

[26] E. J. Sanz-Arigita *et al.*, "Loss of 'Small-World' Networks in Alzheimer's Disease: Graph Analysis of fMRI Resting-State Functional Connectivity," *PLoS One*, vol. 5, no. 11, 2010.

[27] S. J. Teipel *et al.*, "Multicenter stability of resting state fMRI in the detection of Alzheimer's disease and amnestic MCI," *NeuroImage Clin.*, vol. 14, pp. 183–194, 2017.

[28] C. G. Yan, R. C. Craddock, X. N. Zuo, Y. F. Zang, and M. P. Milham, "Standardizing the intrinsic brain: Towards robust measurement of inter-individual variation in 1000 functional connectomes," *Neuroimage*, vol. 80, pp. 246–262, 2013.

[29] R. Gopalan, R. Li, and R. Chellappa, "Unsupervised adaptation across domain shifts by generating intermediate data representations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 11, pp. 2288–2302, 2014.

[30] H. V. Nguyen, H. T. Ho, V. M. Patel, and R. Chellappa, "DASH-N: Joint Hierarchical Domain Adaptation and Feature Learning," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5479–5491, 2015.

[31] V. M. Patel, R. Gopalan, R. Li, and R. Chellappa, "Visual Domain Adaptation: A survey of recent advances," *IEEE Signal Process. Mag.*, vol. 32, no. 3, pp. 53–69, 2015.

[32] A. Van Opbroek, M. A. Ikram, M. W. Vernooij, and M. De Bruijne, "Transfer learning improves supervised image segmentation across imaging protocols," *IEEE Trans. Med. Imaging*, vol. 34, no. 5, pp. 1018–1030, 2015.

[33] S. Conjeti *et al.*, "Supervised domain adaptation of decision forests: Transfer of models trained in vitro for in vivo intravascular ultrasound tissue characterization," *Med. Image Anal.*, vol. 32, pp. 1–17, 2016.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2018.2839771, IEEE Journal of Biomedical and Health Informatics

[34] C. Wachinger and M. Reuter, "Domain adaptation for Alzheimer's disease diagnostics," *Neuroimage*, vol. 139, pp. 470–479, 2016.

[35] M. Goetz *et al.*, "DALSA: Domain adaptation for supervised learning from sparsely annotated MR images," *IEEE Trans. Med. Imaging*, vol. 35, no. 1, pp. 184–196, 2016.

[36] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars, "Unsupervised visual domain adaptation using subspace alignment," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 2960–2967.

[37] W. Koch *et al.*, "Diagnostic power of default mode network resting state fMRI in the detection of Alzheimer's disease," *Neurobiol. Aging*, vol. 33, no. 3, pp. 466–478, 2012.

[38] Y. Guo, T. Hastie, and R. Tibshirani, "Regularized linear discriminant analysis and its application in microarrays.," *Biostatistics*, vol. 8, no. 1, pp. 86–100, 2007.

[39] G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, and E. Stadlan, "Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of the Department of Health and Human Services Task Force on Alzheimer's Disease," *Neurology*, vol. 34, no. 7, pp. 939–944, 1984.

[40] X.-W. Song *et al.*, "REST: a toolkit for resting-state functional magnetic resonance imaging data processing.," *PLoS One*, vol. 6, no. 9, p. e25031, 2011.

[41] Y. Chao-Gan and Z. Yu-Feng, "DPARSF: A MATLAB Toolbox for 'Pipeline' Data Analysis of Resting-State fMRI," *Front Syst Neurosci*, vol. 4, p. 13, 2010.

[42] N. Tzourio-Mazoyer *et al.*, "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain.," *Neuroimage*, vol. 15, no. 1, pp. 273–289, 2002.

[43] M. G. Kendall, "A New Measure of Rank Correlation," *Biometrika*, vol. 30, no. 1/2, pp. 81–93, 1938.

[44] L. L. Zeng *et al.*, "Identifying major depression using whole-brain functional connectivity: A multivariate pattern analysis," *Brain*, vol. 135, no. 5, pp. 1498–1507, 2012.

[45] L. Van Der Maaten and G. Hinton, "Visualizing Data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, 2008.

[46] C. Y. Wee *et al.*, "Identification of MCI individuals using structural and functional connectivity networks," *Neuroimage*, vol. 59, no. 3, pp. 2045–2056, 2012.

[47] M. Xia, J. Wang, and Y. He, "BrainNet Viewer: A Network Visualization Tool for Human Brain Connectomics," *PLoS One*, vol. 8, no. 7, 2013.

[48] B. Jie, D. Zhang, W. Gao, Q. Wang, C. Y. Wee, and D. Shen, "Integration of network topological and connectivity properties for neuroimaging classification," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 2, pp. 576–589, 2014.

[49] M. Dyrba, M. Grothe, T. Kirste, and S. J. Teipel, "Multimodal analysis of functional and structural disconnection in Alzheimer's disease using multiple kernel SVM," *Hum. Brain Mapp.*, vol. 36, no. 6, pp. 2118–2131, 2015.

[50] T. M. Schouten *et al.*, "Combining anatomical, diffusion, and resting state functional magnetic resonance imaging for individual classification of mild and moderate Alzheimer's disease," *NeuroImage Clin.*, vol. 11, pp. 46–51, 2016.

[51] A. Khazaee, A. Ebrahimzadeh, and A. Babajani-Feremi, "Identifying patients with Alzheimer's disease using resting-state fMRI and graph theory," *Clin. Neurophysiol.*, vol. 126, no. 11, pp. 2132–2141, 2015.

[52] T. Heimann, P. Mountney, M. John, and R. Ionasec, "Learning without labeling: Domain adaptation for ultrasound transducer localization," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2013, vol. 8151 LNCS, no. PART 3, pp. 49–56.

[53] T. Schlegl, J. Ofner, and G. Langs, "Unsupervised pre-training across image domains improves lung tissue classification," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2014, vol. 8848, pp. 82–93.

[54] R. Gopalan, R. Li, and R. Chellappa, "Domain adaptation for object recognition: An unsupervised approach," in *Proceedings of the IEEE International Conference on Computer Vision*, 2011, pp. 999–1006.

[55] K. Saenko, B. Kulis, M. Fritz, and T. Darrell, "Adapting visual category models to new domains," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2010, vol. 6314 LNCS, no. PART 4, pp. 213–226.

[56] I. H. Jhuo, D. Liu, D. T. Lee, and S. F. Chang, "Robust visual domain adaptation with low-rank reconstruction," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2168–2175.

[57] S. Shekhar, V. M. Patel, H. V. Nguyen, and R. Chellappa, "Generalized domain-adaptive dictionaries," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2013, pp. 361–368.

[58] J. Hoffman, E. Rodner, J. Donahue, T. Darrell, and K. Saenko, "Efficient Learning of Domain-invariant Image Representations," *Iclr*, pp. 1–9, 2013.

[59] N. C. Fox *et al.*, "Presymptomatic hippocampal atrophy in Alzheimer's disease. A longitudinal MRI study," *Brain*, vol. 119, no. 6, pp. 2001–2007, 1996.

[60] Y. He *et al.*, "Regional coherence changes in the early stages of Alzheimer's disease: A combined structural and resting-state functional MRI study," *Neuroimage*, vol. 35, no. 2, pp. 488–500, 2007.

[61] J. Koppel, S. Sunday, T. E. Goldberg, P. Davies, E. Christen, and B. S. Greenwald, "Psychosis in Alzheimer's disease is associated with frontal metabolic impairment and accelerated decline in working memory: Findings from the Alzheimer's disease neuroimaging initiative," *Am. J. Geriatr. Psychiatry*, vol. 22, no. 7, pp. 698–707, 2014.

[62] Z. Wang *et al.*, "Spatial patterns of intrinsic brain activity in mild cognitive impairment and alzheimer's disease: A resting-state functional MRI study," *Hum. Brain Mapp.*, vol. 32, no. 10, pp. 1720–1740, 2011.

[63] Z. Dai *et al.*, "Discriminative analysis of early Alzheimer's disease using multi-modal imaging and multi-level characterization with multi-classifier (M3)," *Neuroimage*, vol. 59, no. 3, pp. 2187–2195, 2012.

[64] J. S. Damoiseaux, K. E. Prater, B. L. Miller, and M. D. Greicius, "Functional connectivity tracks clinical deterioration in Alzheimer's disease," *Neurobiol. Aging*, vol. 33, no. 4, 2012.

[65] S. M. Kim *et al.*, "Regional cerebral perfusion in patients with Alzheimer's disease and mild cognitive impairment: effect of APOE epsilon4 allele.," *Neuroradiology*, vol. 55, no. 1, pp. 25–34, 2013.

[66] F. Liu *et al.*, "Multivariate classification of social anxiety disorder using whole brain functional connectivity.," *Brain Struct. Funct.*, pp. 101–115, 2013.

[67] H. Zhang *et al.*, "Topographical Information-Based High-Order Functional Connectivity and Its Application in Abnormality Detection for Mild Cognitive Impairment," *J. Alzheimer's Dis.*, vol. 54, no. 3, pp. 1095–1112, 2016.

[68] X. Chen, H. Zhang, L. Zhang, C. Shen, S. W. Lee, and D. Shen, "Extraction of dynamic functional connectivity from brain grey matter and white matter for MCI classification," *Hum. Brain Mapp.*, vol. 38, no. 10, pp. 5019–5034, 2017.

[69] L. Qiao, H. Zhang, M. Kim, S. Teng, L. Zhang, and D. Shen, "Estimating functional brain networks by incorporating a modularity prior," *Neuroimage*, vol. 141, pp. 399–407, 2016.