Published in final edited form as:
Med Image Comput Comput Assist Interv. 2017 September ; 10435: 205–213. doi: 10.1007/978-3-319-66179-7_24.

Personalized Diagnosis for Alzheimer's Disease

Yingying Zhu[1], **Minjeong Kim**[1], **Xiaofeng Zhu**[1], **Jin Yan**[3], **Daniel Kaufer**[2], and **Guorong Wu**[1]

[1]Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

[2]Department of Neurology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

[3]Department of Cancer Biology, Duke University, Durham, NC 27705, USA

Abstract

Current learning-based methods for the diagnosis of Alzheimer's Disease (AD) rely on training a general classifier aiming to recognize abnormal structural alternations from homogenously distributed dataset deriving from a large population. However, due to diverse disease pathology, the real imaging data in routine clinic practices is highly complex and heterogeneous. Hence, prototype methods commonly performing well in the laboratory cannot achieve expected outcome when applied under the real clinic setting. To address this issue, herein we propose a novel personalized model for AD diagnosis. We customize a subject-specific AD classifier for the new testing data by iteratively reweighting the training data to reveal the latent testing data distribution and refining the classifier based on the weighted training data. Furthermore, to improve estimation of diagnosis result and clinical scores at the individual level, we extend our personalized AD diagnosis model to a joint classification and regression scenario. Our model shows improved performance on classification and regression accuracy when applied on Magnetic Resonance Imaging (MRI) selected from Alzheimer's Disease Neuroimaging Initiative (ADNI) database. Our work pin-points the clinical potential of personalized diagnosis framework in AD.

1 Introduction

Alzheimer's Disease is one of the most common neurodegenerative disorders, which leads to gradual progressive memory loss, cognition declines, loss of functional abilities, and ultimate death [1–4]. Modern imaging technique MRI offers a non-invasive way to observe the abnormal structure changes of AD progression *in vivo*. In order to facilitate the MRI diagnosis of AD, a number of machine learning approaches have been developed to recognize AD-related altered brain structure [5, 6].

Most current learning-based methods train a general classifier (such as kernel Support Vector Machine) to find a hyper-plane to separate two groups in a high dimensional non-linear space, which is suitable for data of homogeneous distribution (as shown in (a)). However, it is evident that AD pathology is of heterogeneous characteristic. Therefore, real

Correspondence to: Yingying Zhu.

data distribution could be too complex to be represented by only one general model. As shown in (b), the real data distribution might have multiple sub-groups of distinct disease patterns due to inter-personal variance (different colors indicate different sub-groups). The incapability of a general classifier to fit heterogeneous data for precise classification prompts us to construct a specific classifier for each sub-group/person to achieve accurate classification. As shown in (b), the green, orange, and pink curves in are ideal classifiers for the green dataset, orange dataset, and pink dataset, respectively. Based on this strategy, we developed a personalized classification model specifically for each testing person desirable for real clinical applications.

The key to the personalized model is to reweight the training data to fit testing data distribution and train a person-specific classifier using the weighted training data. Figure 1(b) and (c) show a toy example on how to construct a personalized classifier. There are three sub-groups of different distributions in the training dataset (green, orange, and pink) and one testing dataset (purple). Although it seems that the center of the testing dataset is closer to the green dataset than the orange dataset, only the distribution of the orange dataset resembles the testing dataset. Therefore, the orange dataset is associated with high-value weights and others are assigned with small value weights (see Fig. 1(c), size is proportional to weights). This weighted dataset reveals the latent testing data distribution and a personalized classifier (purple curve in Fig. 1(c)) can be learned from the weighted training dataset. Furthermore, to optimize weights for building the classifier, we developed an integrated solution which makes learning the training data weights and training the personalized classifier simultaneous.

This personalized training strategy can sort out the data heterogeneity issue and produce more accurate diagnosis results at the individual level. Moreover, we extend the personalized diagnosis model to a joint classification (on the binary clinical labels) and regression (on the continuous clinical scores) such that we can further improve the accuracy of diagnosis by utilizing both imaging and phenotype data. We evaluated the proposed personalized AD diagnosis model on ADNI dataset and achieved more than **8%** improvement on average in identifying AD and MCI (Mild Cognition Impairment) subjects, compared to using general classification models.

2 Methods

2.1 Generalized Classification Model

Suppose the training set $\boldsymbol{X}$ consist of N subjects, denoted by $\boldsymbol{X} = \{\boldsymbol{x}_i | i = 1, \cdots, N\}$, where each $\boldsymbol{x}_i$ is the feature vector extracted from MRI. Each training subject also has a clinical label l_i to identify whether the underlying subject is at MCI stage ('−1') or has converted to AD ('+1'). These clinical labels form a set $\boldsymbol{L} = \{l_i | i = 1, \ldots, N\}$. Hereafter, we take the kernel SVM as the example of the generalized classification model to illustrate the idea of our personalized AD diagnosis. Kernel SVM seeks to learn a non-linear mapping $\boldsymbol{\varphi}$ to determine the label for the new testing data $\boldsymbol{y}$, where $\boldsymbol{\varphi}$ is essentially the weighted kernel distances with respect to each known instance $\boldsymbol{x}_i$ in the training dataset $\mathbf{X}$, i.e., $\varphi(\boldsymbol{y}) = \sum_{j=1}^{N} \alpha_i k_C(\boldsymbol{x}_i, \boldsymbol{y})$, where $k_C(\boldsymbol{x}_i, \boldsymbol{y})$ denotes the kernel distance from $\boldsymbol{x}_i$ to $\boldsymbol{y}$ in the high dimensional non-linear space. The classification coefficient $\boldsymbol{\alpha} = \{\alpha_i | i = 1, \ldots, N\}$ can be optimized from the training dataset $\boldsymbol{X}$ via the following classic energy function:

$$\arg\ \min_{\boldsymbol{\alpha}} \frac{1}{2}\boldsymbol{\alpha}^T \boldsymbol{K}_C \boldsymbol{\alpha} + \mu E_C(\varphi)\,. \quad (1)$$

The first term is the regularization term where $\boldsymbol{K}_c$ is a $N \times N$ kernel matrix with each element $k_C(\boldsymbol{x}_i, \boldsymbol{x}_j)$ measuring the distance between any two training data $\boldsymbol{x}_i$ to $\boldsymbol{x}_j$ ($i, j = 1, \ldots, N$) in the high dimensional non-linear space. The second term is the misclassification error term $E_C(\varphi) = \sum_{i=1}^{N} \|\varphi(\boldsymbol{x}_i) - l_i\|_h$, where $\|v\|_h = \max(0, v)$ is the hinge loss function. μ is a scalar balancing the regularization term and misclassification error term. The classification coefficient $\boldsymbol{\alpha}$ is optimized to fit the whole population in Eq. (1). If the data distribution is as homogeneous as the example shown in (a), a generalized classifier can achieve good performance on classification. However, real clinical data usually has complex distribution due to heterogeneous characteristics of AD pathology. Hence, one general classifier alone is not sufficient to cover all individuals.

2.2 Personalized Classification Model

To address the issue of heterogeneity, we propose to learn a person-specific classifier by leveraging the most relevant data in the training dataset. The training data are reweighted to reveal the testing subject distribution. Compared to the general model, we measure the relevance degree of each training data $\boldsymbol{x}_i$ w.r.t. the new testing data $\boldsymbol{y}$ denoted by $\boldsymbol{r} = \{r_i\}_{i=1,\ldots,N}$. In contrast to the general model which treats all training data uniformly, here, we penalize the misclassification error for different training data w.r.t the relevance degree to testing data. To achieve personalized AD diagnosis, we adjust the energy function of general classification model to personalized classification model in three ways:

1. The misclassification errors are weighted based on the relevance degree $\boldsymbol{r}$, it turns to the weighted average across all training data: $E_c(\varphi, \boldsymbol{r}) = \sum_{i=1}^{N} r_i \|\varphi(\boldsymbol{x}_i) - l_i\|_h^2$.

2. The insight of personalized classifier is to re-weight each training data such that the difference between the distribution of testing data and weighted training data distribution is minimized. Therefore, we introduce the additional distribution mismatch term which is related to the relevance values $\boldsymbol{r}$. First, we need to estimate the distribution for testing data $\boldsymbol{y}$. Recall that the challenge in medical imaging area is the limited number of data with label information. Actually, it is not difficult to find a sufficient number of unlabeled data. Therefore, we propose to construct the distribution for the testing data $\boldsymbol{y}$ from another unlabeled dataset, denoted by $\boldsymbol{U}$ ($\boldsymbol{U} \cap \boldsymbol{X} = \emptyset$). Specifically, we select the top similar M data to $\boldsymbol{y}$ from $\boldsymbol{U}$ to form the testing dataset $\boldsymbol{Y} = \{\boldsymbol{y}_j | \boldsymbol{y}_j \in \boldsymbol{U}, j = 1, \ldots, M\}$ (we set $M = 4$ in our experiment). Since the size of testing dataset is very small ($M \ll N$), it is hard to calculate the data density in characterizing the distribution of $\boldsymbol{Y}$. To avoid the unreliable estimation of data density, we resort to the Kernel Mean Matching

(KMM) method [7], which is able to measure the distribution dissimilarity in the high dimensional or even infinite dimensional Reproducing Kernel Hilbert Space (RKHS) H. Specifically, we define the distribution mismatch term $\mathscr{D}(\mathbf{X},\mathbf{Y},\boldsymbol{r})$ as the function of $\boldsymbol{r}$ by:

$$\mathscr{D}(\mathbf{X},\mathbf{Y},\boldsymbol{r}) = \left\| \frac{1}{N}\sum_{i=1}^{N} r_i \phi(\boldsymbol{x}_i) - \frac{1}{M}\sum_{j=1}^{M} \phi(\boldsymbol{y}_j) \right\|_2^2, \quad (2)$$

where $\boldsymbol{\phi}$ is a non-linear mapping from image feature space to the RKHS. The intuition behind Eq. (2) is to adjust the distribution of training dataset $\boldsymbol{X}$ based on the relevance degrees such that the weighted distribution of training dataset (first term in Eq. (2)) can fit the distribution of testing dataset $\boldsymbol{Y}$, i.e., $\boldsymbol{X}$ and $\boldsymbol{Y}$ are comparable in the RKHS. To make Eq. (2) solvable, “kernel trick” is used to compute the pairwise kernel distance in H, rather than the exact value $\boldsymbol{\phi}(\boldsymbol{x}_i)$. Thus, we use $\mathbf{\Pi} = [\boldsymbol{\pi}_{ij}]_{N\times N}\,(i, j = 1, \cdots, N)$ to denote the $N \times N$ kernel matrix where each element $\boldsymbol{\pi}_{ij} = k_D(\boldsymbol{x}_i, \boldsymbol{x}_j)$ measures the kernel distance between training data $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ in RKHS H. Similarly, we use $\boldsymbol{\theta} = [\theta_i]_{i=1,\ldots,N}$ to denote a N-length column vector where the i-th element $\theta_i = \frac{N}{M}\sum_{j=1}^{M} k_D\left(\boldsymbol{x}_i, \boldsymbol{y}_j\right)$ measures the average distance of the training data $\boldsymbol{x}_i$ to all testing data. It is worth noting that the kernel function k_D used here might be different with the kernel function k_c in kernel SVM (Eq. (1)). After that, we turn Eq. (2) into a quadratic term as:

$$E_D(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{r}) = \frac{1}{2}(\boldsymbol{r})^T \mathbf{\Pi}\boldsymbol{r} - \boldsymbol{\theta}^T\boldsymbol{r}, \quad (3)$$

3. Since the training error term $E_c(\boldsymbol{\varphi}, \boldsymbol{r})$ and the distribution mismatching term $E_D(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{r})$ are both influenced by the relevance degrees $\boldsymbol{r}$, we can leverage the $\boldsymbol{r}$ to jointly minimize the classification error and distribution difference. By doing so, we can guarantee that the relevance degrees are optimized towards the eventual goal of improving the classification accuracy for the new testing data.

Overall energy function—By integrating the above three modifications, the overall energy function for personalized AD diagnosis can be defined as:

$$\arg\ \min_{\boldsymbol{\alpha},\boldsymbol{r}} \frac{1}{2}\boldsymbol{\alpha}^T \boldsymbol{K}_C \boldsymbol{\alpha} + \mu E_C(\varphi, \boldsymbol{r}) + \lambda E_D(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{r}) \quad (4)$$

where λ is the scalar used to control the strength of distribution matching. Since the clinic label of the testing data is unknown yet, the estimation of the weights for training data is driven by the distribution mismatch term $E_D(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{r})$ in an unsupervised manner. In addition, the estimation of relevance degrees $\boldsymbol{r}$ are jointly driven by minimizing the classification error and distribution mismatch.

Optimization—Equation (4) is a bi-convex quadratic problem [8], i.e., $E(\boldsymbol{\alpha})$ is convex if $\boldsymbol{r}$ is fixed and $E(\boldsymbol{r})$ is convex if $\boldsymbol{\alpha}$ is fixed. Under these conditions, an alternate gradient search approach is guaranteed to monotonically decrease the objective function. Hence, we alternatively optimize Eq. (4) w. r. t. $\boldsymbol{r}$ and $\boldsymbol{\alpha}$ until converge [9].

Discussion—Conventional subject selection approaches, however, are usually separately performed prior to train the classifier, and thus resulting in a sequential two-step strategy. Therefore, the selected training data in the two-step strategy might not be optimal for classification, since there is no chance to refine the subject selection procedure. Since the personalized classifier is free of the less relevant training samples, the customized mapping function in our personalized classifier is more robust to the new testing subject than the general classifier learned from entire training data, as demonstrated in (c).

2.3 Advance Personalized AD Diagnosis Model

In many medical applications, clinical scores such as Mini-Mental State Examination (MMSE) and Clinical Dementia Rating (CDR) scores are widely used to quantify memory loss and behavior abnormality and facilitate the clinical AD diagnosis. Since the clinical scores have higher correlations than the imaging features, there is an increasing trend to integrate classification (for the binary diagnosis labels) and regression (for the continuous clinical scores). Hence, we go one step further to present the advanced personalized AD diagnosis model. Suppose each training data $\boldsymbol{x}_i$ has the clinical scores c_i which forms the set of clinical scores $\mathbf{C} = \{c_i | i = 1, \ldots, N\}$. For consistency, we use the kernel Support Vector Regression (SVR) model to learn another non-linear mapping function $\psi(\boldsymbol{y}) = \sum_{i=1}^{N} \beta_i k_R(\boldsymbol{x}_i, \boldsymbol{y})$ to determine the scores for the new testing data $\boldsymbol{y}$ based on the weighted average of kernel distance $k_R(\boldsymbol{x}_i, \boldsymbol{y})$ to all training data. The regression coefficients $\boldsymbol{\beta} = \{\beta_i | i = 1, \ldots, N\}$ can be optimized by:

$$\arg\ \min_{\boldsymbol{\beta}} \frac{1}{2}\boldsymbol{\beta}^T \boldsymbol{K}_R \boldsymbol{\beta} + \eta E_R(\psi), \quad (5)$$

where η is the scalar balancing the regularization term and regression error term $E_R(\psi) = \sum_{i=1}^{N} \|\psi(x_i) - c_i\|_2^2$. $\boldsymbol{K}_R$ is the $N \times N$ kernel matrix with each element $k_R(x_i, x_j)$ measuring the kernel distance between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ in the regression problem. To personalize the regression, we turn the regression error term $E_R(\psi)$ into the weighted average across all training data as $E_R(\psi, \boldsymbol{r}) = \sum_{i=1}^{N} r_i \|\psi(\boldsymbol{x}_i) - c_i\|_2^2$. Thus, the energy function of personalized regression can be derived as:

$$\arg\ \min_{\boldsymbol{\beta}, \boldsymbol{r}} \frac{1}{2}\boldsymbol{\beta}^T \boldsymbol{K}_R \boldsymbol{\beta} + \mu E_C(\psi, \boldsymbol{r}) + \lambda E_D(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{r}), \quad (6)$$

Furthermore, we integrate personalized classification and regression and derive the overall energy function of advanced personalized AD diagnosis model as:

$$\arg\ \min_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{r}} \frac{1}{2}\boldsymbol{\alpha}^T\boldsymbol{K}_C\boldsymbol{\alpha} + \frac{1}{2}\boldsymbol{\beta}^T\boldsymbol{K}_R\boldsymbol{\beta} + \mu E_C(\varphi, \boldsymbol{r}) + \eta E_R(\psi, \boldsymbol{r}) + \lambda E_D(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{r})\,. \quad (7)$$

Equation (7) can be similarity solved by alternatively updating $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{r}$ until converge.

3 Experiments

We evaluate our proposed personalized AD diagnosis model on 150 MCI and 150 AD subjects selected from ADNI database, each has the clinical scores such as ADAS-COG (Alzheimer's Disease Assessment Scale-Cognitive Subscale) and MMSE. For each subject, we first segment each image into white matter (WM), gray matter (GM), cerebral spinal fluid (CSF). Then, we register the AAL template with 90 manually labeled ROIs (regions of interest) to the underlying subject image. We concatenate the tissue percentiles across 90 ROIs as the morphological feature for each subject.

The accuracy of classification and regression is evaluated by leave-one-out strategy. For each leave-one-out case, we divide the remaining subjects into five folds. One-fold data is used as the validation dataset for parameter turning, one-fold is used as the candidature dataset for augmenting testing data and the left three-fold data are use as the training dataset. The optimal parameters are learned by an exhaustive strategy on the validation dataset. The search range for parameters is set to $[10^{-4}, 10^{4}]$. Three statistical measures are used to evaluate classification, including accuracy (ACC), sensitivity (SEN) and specificity (SPEC). Root Mean Square Error (RMSE) and Correlation Coefficients (CC) are used to evaluate the regression performance for two popular clinical measurements, i.e., ADAS-Cog and MMSE.

Kernel SVM/SVR(generalized model) are the baseline methods in comparison. In the following experiments, we evaluate the classification and regression separately. We use the RBF kernel and the regularization parameters are tuned using five-fold inner cross-validation. In classification, we compare our personalized classifier (Eq. 4) with kernel SVM and convention subject selection (estimate the weight based on feature similarity) followed by kernel SVM called SS+SVM. Similarly, we evaluate the regression performance for the baseline kernel SVR, SS+SVR, and our personalized regressor. Our advance personalized AD diagnosis model (called Personalized SVM+SVR) is both evaluated in classification and regression tasks, in order to show the benefit from joint classification and regression.

Evaluation of classification performance—The ACC, SEN, and SPEC results in identifying MCI/AD and NC/MCI/AD subjects by kernel SVM, SS+SVM, personalized SVM, and personalized SVM+SVR are shown in Fig. 2(a). In general, all personalized approaches achieve higher accuracy than the baseline kernel SVM method which does not have any adjustment to the testing subject. Since our personalized SVM method can jointly select the most relevant subjects and train the classifier, it achieves overall 1.3% improvement in terms of ACC value over the naïve SS+SVM approach that selects training data and trains the classifier separately. Furthermore, our personalized SVM+SVR can obtain additional 3.1% improvement of ACC value over the personalized SVM method, which shows the substantial benefit of joint classification and regression. It is worth noting that our advanced personalized model (personalized SVM+SVR) achieves 8.3% improvement in terms of ACC value, compared to generalized model (kernel SVM).

Evaluation of regression performance—Since the clinical scores of each testing subject are known, we calculate the RMSE and CC values between the ground truth and the estimated score by four competing methods. We show the MMSE and CC results of estimating MMSE and ADAS-Cog measurements by kernel SVR, SS+SVR, personalized SVR, and personalized SVM+SVR in Fig. 2(b) and (c), respectively. It is apparent that (1) all personalized method beat the generalized regression method (kernel SVR); (2) our personalized SVR outperforms the naïve SS+SVR method due to the advantage of joint weighting training data and regression; (3) personalized SVM+SVR has the minimal MMSE value and large CC value between the ground truth and estimated scores, indicate the advantage of allowing clinical labels to guide the regression of clinical scores.

Evaluation of personalized diagnosis model with respect to data heterogeneity—Since the main objective of our study is to address the heterogeneous issue of imaging data by using personalized model, we specifically evaluate the performance of the personalized model with respect to the heterogeneity in the observed imaging data. Here we assume that the data heterogeneity proportionally increases as the size of imaging data becomes larger and larger. Therefore, we examine the performance of classification and regression w.r.t. the different number of the training data, as shown in Fig. 3. We run the two-sample t-test for results in Fig. 3 and find the improvement is significant with $p < 0.05$. One can observe that (1) both general and personalized models perform at similar accuracy when the size of training data is small; (2) as the number of training subjects increases, the personalized model achieved much higher performance accuracy compared to general model in terms of both classification and regression tasks, although the performance accuracy of all methods increase consistently; (3) the improvement of personalized model over the generalized model becomes more prominent as the subject number increases. These results prove that the personalized model we propose is superior to the general model when applied on a large heterogeneous dataset and thus is of potential for clinical practice.

4 Conclusion

To address the heterogeneous issue of image-based diagnosis of AD, we construct a personalized diagnosis model in this work. In this model, we establish a subject-specific AD classifier by reweighting training data to reveal the latent distribution for each testing data and simultaneously refining classifier. We further improve the diagnosis performance at the individual level by establishing a joint classification and regression scenario. Finally, we evaluate our method on ADNI dataset for both clinical label and clinical score estimation compared to the state-of-art counterpart methods and demonstrate the potential of our personalized model in translating computer assisted diagnosis method into routine clinical practice.

References

1. Viola K, et al. Towards non-invasive diagnostic imaging of early-stage Alzheimer's disease. Nat Nanotechnol. 2015; 10:91–98. [PubMed: 25531084]
2. Thompson PM, Hayashi KM, Dutton RA, Chiang M-C, Leow AD, Sowell ER, et al. Tracking Alzheimer's disease. Annals of New York Academy of Sciences. 2007; 1097:198–214.
3. Zhu Y, , Zhu X, , Kim M, , Shen D, , Wu G. Early diagnosis of alzheimer's disease by joint feature selection and classification on temporally structured support vector machine. In: Ourselin S, Joskowicz L, Sabuncu Mert R, Unal G, , Wells W, editorsMICCAI 2016 LNCS Vol. 9900. Springer; Cham: 2016 264272
4. Wang Z, , Zhu X, , Adeli E, , Zhu Y, , Zu C, , Nie F, , Shen D, , Wu G. Progressive graph-based transductive learning for multi-modal classification of brain disorder disease. In: Ourselin S, Joskowicz L, Sabuncu Mert R, Unal G, , Wells W, editorsMICCAI 2016 LNCS Vol. 9900. Springer; Cham: 2016 291299
5. Lindberg O, et al. Hippocampal shape analysis in Alzheimer's disease and frontotemporal lobar degeneration subtypes. J Alzheimers Dis. 2012; 30:355–365. [PubMed: 22414571]
6. Pettigrew C, et al. Cortical thickness in relation to clinical symptom onset in preclinical AD. Neuroimage: Clinical. 2016; 15:116–122.
7. Gretton A, et al. Covariate shift by kernel mean matching. Dataset Shift in Machine Learning. 2009:123–135.
8. Boyd S, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends in Machine Learning. 2011; 3:1–122.
9. Zhu Y, Lucey S. Convolutional sparse coding for trajectory reconstruction. TPAMI. 2015; 37:529–540.

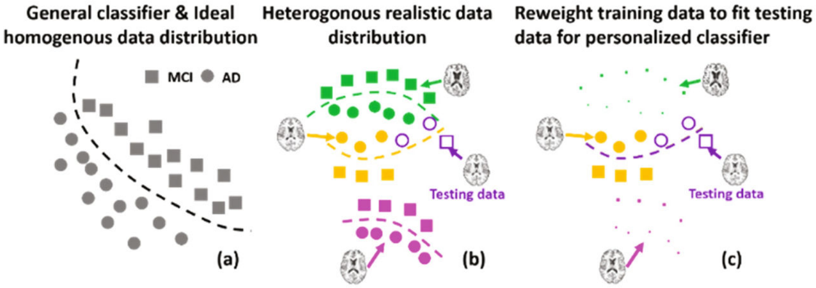
General classifier & Ideal homogenous data distribution
Heterogonous realistic data distribution
Reweight training data to fit testing data for personalized classifier
MCI
AD
Testing data
Testing data
(a)
(b)
(c)

Fig. 1.

(a) Conventional methods with a general classifier applied on homogeneously distributed data; (b) Heterogeneously distributed realistic data requiring person-specific classification solution; (c) Proposed personalized classification model by re-weighting training data to fit testing data distribution. (**Best viewed in color**)

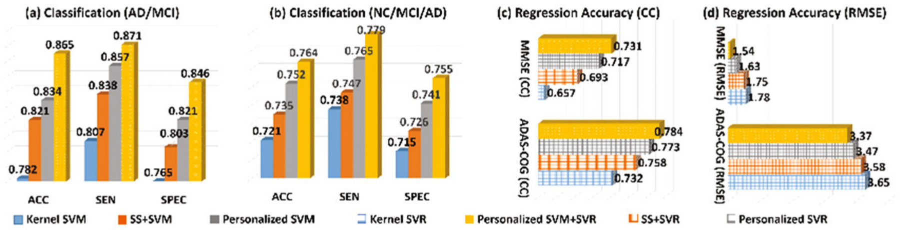

Fig. 2.

Classification performance in identifying MCI/AD, NC/MCI/AD subjects and regression performance in estimating MMSE and ADAS-Cog scores. (CC represent correlation coefficients and RMSE represents root mean square error. **Best viewed in color**)

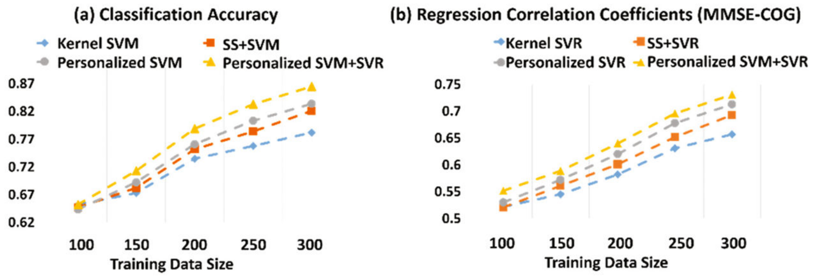

Fig. 3.

The performance of personalized model vs. general model with respect to different number of training data. (**Best viewed in color**)