Biomarker Extraction Based on Subspace Learning for the Prediction of Mild Cognitive Impairment Conversion

Ying Li,1,2 Yixian Fang,3 Jiankun Wang,4 Huaxiang Zhang,2 and Bin Hu2,5

1Key Laboratory of TCM Data Cloud Service in Universities of Shandong, Shandong Management University, Jinan 250357, China
2School of Information Science and Engineering, Shandong Normal University, Jinan 250358, China
3School of Mathematics and Statistics, Qilu University of Technology, Jinan 250353, China
4Shandong Big Data Center, Jinan 250011, China
5Key Laboratory of Wearable Computing of Gansu Province, Lanzhou University, Lanzhou 730000, China

Correspondence should be addressed to Huaxiang Zhang; huaxzhang@163.com and Bin Hu; bh@lzu.edu.cn

Received 3 February 2021; Accepted 13 August 2021; Published 2 September 2021

Academic Editor: Lei Zhang

Copyright © 2021 Ying Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Accurate recognition of progressive mild cognitive impairment (MCI) is helpful to reduce the risk of developing Alzheimer’s disease (AD). However, it is still challenging to extract effective biomarkers from multivariate brain structural magnetic resonance imaging (MRI) features to accurately differentiate the progressive MCI from stable MCI. We develop novel biomarkers by combining subspace learning methods with the information of AD as well as normal control (NC) subjects for the prediction of MCI conversion using multivariate structural MRI data. Specifically, we first learn two projection matrices to map multivariate structural MRI data into a common label subspace for AD and NC subjects, where the original data structure and the one-to-one correspondence between multiple variables are kept as much as possible. Afterwards, the multivariate structural MRI features of MCI subjects are mapped into a common subspace according to the projection matrices. We then perform the self-weighted operation and weighted fusion on the features in common subspace to extract the novel biomarkers for MCI subjects. The proposed biomarkers are tested on Alzheimer’s Disease Neuroimaging Initiative (ADNI) dataset. Experimental results indicate that our proposed biomarkers outperform the competing biomarkers on the discrimination between progressive MCI and stable MCI. And the improvement from the proposed biomarkers is not limited to a particular classifier. Moreover, the results also confirm that the information of AD and NC subjects is conducive to predicting conversion from MCI to AD. In conclusion, we find a good representation of brain features from high-dimensional MRI data, which exhibits promising performance for predicting conversion from MCI to AD.

1. Introduction

Alzheimer’s disease (AD) characterized by memory loss and cognitive decline is the most prevalent neurodegenerative disease [1, 2]. Mild cognitive impairment (MCI) is regarded as the prodromal stage of AD with possibility to develop AD. Individuals with MCI can carry out daily activities, but their thinking abilities have mild and measurable changes [3]. On average, 32 percent of individuals with MCI will convert to AD within 5 years [4]. Therefore, it is critical to identify MCI as early as possible, so that we can delay the progress of AD by the well-targeted treatment. The development of neuroimaging techniques provides powerful tools for early prediction of AD. Structural magnetic resonance imaging (MRI) with high spatial resolution, high availability, noninvasive nature, and moderate costs is an extensively used neuroimaging modality. Numerous structural MRI-based biomarkers have been extracted for the AD detection at different stages [5–13]. For instance, in [6], spatial frequency components of cortical thickness were used for individual AD identification based on incremental learning. In [13], an individual network was constructed using six types of
morphological features to improve the accuracy of AD and MCI diagnoses. However, since the pathological variations are subtle at the MCI stage, it is still challenging to develop more advanced biomarkers to accurately predict the conversion from MCI to AD.

According to whether the MCI subjects will convert to AD or not within a given time period (i.e., 3 years), they are separated into two categories: progressive MCI (pMCI) and stable MCI (sMCI). Previous studies [14, 15] have shown that the subjects with pMCI are similar to AD while subjects with sMCI are more like normal control (NC). As a result, the classification between AD and NC is a simple version of that between pMCI and sMCI. Due to the high heterogeneity of MCI population, it is effective to take advantage of AD and NC information in MCI conversion prediction, such as feature selection and classifier training. Studies [14–22] also have demonstrated that the information of AD and NC subjects is helpful in distinguishing pMCI subjects from sMCI subjects. In [16, 17], the data of AD and NC subjects was used to build classifier for the discrimination between pMCI and sMCI subjects. In [18–20], the AD and NC subjects were regarded as labeled samples while MCI subjects were taken as unlabeled samples, and a semi-supervised learning approach was applied to dividing MCI subjects into normal-like and AD-like categories. In [14], to distinguish pMCI from sMCI, a semisupervised low-density separation (LDS) method was used to integrate AD and NC information. In [21], a novel domain transfer learning method drawing support from AD and NC subjects was used for MCI conversion prediction. Besides, some studies extracted novel biomarkers for MCI conversion prediction by information propagation from AD and NC subjects to MCI subjects. For instance, in [22], the information was propagated from AD and NC subjects to MCI subjects by a weighting function, and the average grading value was computed for MCI classification. In [15], the disease labels of AD and NC subjects were propagated to MCI subjects using elastic net technique, and a global grading biomarker was developed.

Owing to the high dimensionality of MRI features, it is difficult to find a good representation of brain features to reveal their subtle pathological variations for MCI conversion prediction [23]. The subspace learning method as a dimension reduction approach has become a hot topic in many fields [24–30]. In the field of AD diagnosis, several subspace learning methods, such as canonical correlation analysis (CCA) [31, 32], independent component analysis (ICA) [33, 34], partial least squares (PLS) [35, 36], locality preserving projection (LPP) [37, 38], linear discriminant analysis (LDA) [38, 39], and locally linear embedding (LLE) [23, 40], have demonstrated promising performance. For instance, in [23], multivariate MRI data were transformed into a locally linear space by LLE algorithm, and the embedded features were used to predict the conversion from MCI to AD. In [34], the risk factors associated with MCI conversion were investigated by combining ICA with the multivariate Cox proportional hazards regression model. In [38], a sparse least square regression framework with LDA and LPP was proposed for feature selection in AD diagnosis. The experimental results verified that subspace learning methods outperformed feature selection methods. Although many subspace learning methods have been applied to the early detection of AD, it is still a challenging problem to map MRI data into a low-dimensional subspace and find representative brain features for detecting the differences between pMCI and sMCI. In addition, it is interesting to investigate how the AD and NC data can provide auxiliary information in this procedure and enhance the performance of MCI classification.

In this work, we propose a method to extract biomarkers of MCI subjects based on subspace learning for predicting conversion from MCI to AD. Specifically, we first learn two projection matrices to map multivariate MRI data of regional cortical thickness (CT) and cortical volume (CV) into a common label subspace with lower dimensions for AD and NC subjects, where the correlation of multiple variables and the original data structure are kept as much as possible. We then use the projection matrices to map the CT and CV data of the MCI subjects into the common subspace to obtain the CT- and CV-based features for MCI subjects accordingly. After that, we perform self-weighted operation and weighted fusion on the CT- and CV-based features in common subspace and extract the novel biomarkers for MCI subjects.

2. Materials and Method

2.1. Image Data and Preprocessing. Data used in this work are acquired from Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu/). We use baseline MRI scans (1.5 T, 1.25 mm × 1.25 mm in-plane spatial resolution, 1.2 mm thick slices) of 528 subjects, which include 142 AD subjects, 163 NC subjects, and 221 MCI subjects. Moreover, the 221 MCI subjects contain 126 pMCI and 95 sMCI subjects. The characteristics of the participants are shown in Table 1.

The image preprocessing involves the following steps: motion correction, nonbrain tissue removal, coordinate transformation, gray matter (GM) segmentation, and reconstruction of GM/white matter boundaries [41–43]. We conducted all preprocessing steps by FreeSurfer v5.3.0 (http://surfer.nmr.mgh.harvard.edu). The reconstruction and segmentation errors are visually checked using FreeView software and manually corrected. After that, surface inflation and registration are performed, followed by cortical thickness and volume measurement calculation [44]. Finally, the images were smoothed by a 30 mm full width at half maximum Gaussian kernel [45]. The images are segmented into 90 regions in the light of the automated anatomical labeling atlas [46], and then, 12 subcortical regions are removed owing to the lack of the thickness features. The average cortical thickness and cortical volume of each region are calculated and used as features.

2.2. Method. Schematic representation of our proposed method is provided in Figure 1. The method includes three steps: (1) Taking AD and NC subjects as auxiliary data, we learn two projection matrices. (2) The MCI subjects are
mapped into subspace according to the projection matrices. 
(3) Self-weighted operation and weighted fusion are performed on the features in the subspace, and the biomarkers are extracted.

2.2.1. Learning Projection Matrices Using Auxiliary Data. In this subsection, with AD and NC subjects as auxiliary data, we learn two projection matrices to map multivariate structural MRI data of regional cortical thickness and volume into a common label subspace, where the original data structure and the one-to-one correspondence between multiple variables are kept as much as possible. Let \( X_{CT} = [x_{1}^{CT}, x_{2}^{CT}, \ldots, x_{n}^{CT}] \in \mathbb{R}^{d \times n} \) and \( X_{CV} = [x_{1}^{CV}, x_{2}^{CV}, \ldots, x_{n}^{CV}] \in \mathbb{R}^{d \times n} \) denote the cortical thickness and cortical volume feature matrices, respectively, where \( n \) is the number of AD and NC subjects, and \( d \) is the number of feature dimensions. Let \( Y \in \mathbb{R}^{n \times c} \) represent a class indicator matrix with 0-1 encoding, where \( c \) is the number of classes. To learn the two projection matrices \( U_{d}^{CT} \) and \( V_{d}^{CV} \), the objective function is defined as follows:

\[
\min_{U,V} Q(U, V) = \lambda I(U, V) + (1 - \lambda) f(U, V) + \alpha g(U, V) + \beta r(U, V). 
\]  

(1)

The first term \( I(U, V) \) is the linear regression from the feature space to the label space, and it guarantees that samples are close to their labels after projection. \( I(U, V) \) is expressed as follows:

\[
I(U, V) = \| Y - X_{CT}^{T} U \|^2_F + \| Y - X_{CV}^{T} V \|^2_F. 
\]  

(2)

The second term maintains the correlation between the CT features and CV features of the same image. It is well known that different morphological features of the same image reflect the same label information from different views. They should be close to each other after projection. Therefore, \( f(U, V) \) is defined as follows:

\[
f(U, V) = \| X_{CT}^{T} U - X_{CV}^{T} V \|^2_F. 
\]  

(3)

The third term \( g(U, V) \) is the graph regularization term, which is used to better exploit the local structural information of the data. We aim to preserve the neighborhood relationship between samples of single morphological feature. Here, we first introduce the graph regularization term for cortical thickness feature \( X_{CT} \). We define an undirected and symmetric graph \( G_{CT} = (V_{CT}, W_{CT}) \), where \( V_{CT} \) is a collection of samples in \( X_{CT} \) and \( W_{CT} \) represents the relations between samples. Each element \( w_{ij}^{CT} \) in \( W_{CT} \) is defined as follows:

\[
w_{ij}^{CT} = \begin{cases} 
\exp \left( -\frac{(x_{i}^{CT} - x_{j}^{CT})^2}{2\sigma^2} \right), & \text{if } x_{i}^{CT} \in N_{k}(x_{j}^{CT}), i \neq j, \\
0, & \text{otherwise}, 
\end{cases} 
\]  

(4)

where \( N_{k}(x_{i}^{CT}) \) denotes the \( k \)-nearest neighbors of \( x_{i}^{CT} \). Let \( a_{i,j} \) denote the \( i \)-th column of \( U^{T}X_{CT} \); then, the graph regularization term for cortical thickness data is formulated as follows:

\[
g_{CT}(U) = \frac{1}{2} \sum_{i,j=1}^{n} \| a_{i} - a_{j} \|^2 || w_{ij}^{CT} = tr(U^{T}X_{CT}L_{CT}^{T}X_{CT}^{T}U), 
\]  

(5)

where \( L_{CT} = D_{CT} - W_{CT} \) is the graph Laplacian matrix and \( D_{CT} \in \mathbb{R}^{n \times n} \) is a diagonal matrix with its diagonal elements \( D_{CT}^{ij} = \sum_{k=1}^{n} w_{ik}^{CT} \).

Similarly, for the cortical volume data \( X_{CV} \), let \( b_{i,j} \) denote the \( i \)-th column of \( V^{T}X_{CV} \). The graph regularization term of volume data is formulated as follows:

\[
g_{CV}(V) = \frac{1}{2} \sum_{i,j=1}^{n} \| b_{i} - b_{j} \|^2 || w_{ij}^{CV} = tr(V^{T}X_{CV}L_{CV}^{T}X_{CV}^{T}V), 
\]  

(6)

where \( w_{ij}^{CV} \) and \( L_{CV} \) are defined as before. The final representation of the graph regularization term is then given by the following:

\[
g(U, V) = g_{CT}(U) + g_{CV}(V) = tr(U^{T}X_{CT}L_{CT}^{T}X_{CT}^{T}U) \\
+ tr(V^{T}X_{CV}L_{CV}^{T}X_{CV}^{T}V). 
\]  

(7)

The last term \( r(U, V) \) controls the scale of projection matrices and avoids overfitting:

\[
r(U, V) = \| U \|_F^2 + \| V \|_F^2. 
\]  

(8)

Besides, \( \lambda, \alpha, \) and \( \beta \) are the three balancing parameters. Based on Equations (2), (3), (7), and (8), we can obtain the

Table 1: Characteristics of the subjects.

<table>
<thead>
<tr>
<th>Variables</th>
<th>NC</th>
<th>sMCI</th>
<th>MCI</th>
<th>pMCI</th>
<th>AD</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of subjects (male/female)</td>
<td>165 (78/87)</td>
<td>95 (63/32)</td>
<td>126 (73/53)</td>
<td>142 (72/70)</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>76.40 ± 5.37</td>
<td>74.94 ± 7.32</td>
<td>73.40 ± 9.25</td>
<td>76.10 ± 7.51</td>
<td></td>
</tr>
<tr>
<td>CDR</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5/1</td>
</tr>
<tr>
<td>MMSE</td>
<td>29.19 ± 0.96</td>
<td>27.69 ± 1.73</td>
<td>26.49 ± 1.70</td>
<td>23.20 ± 2.01</td>
<td></td>
</tr>
</tbody>
</table>

CDR: Clinical Dementia Rating; MMSE: Mini-Mental State Examination.
Projection matrices learning based on AD and NC subjects

Linear regression

CT space

Local structure preservation

Correlation analysis

CV space

Local structure preservation

Linear regression

Projection matrix $U$ learning

Projection matrix $V$ learning

Common subspace

Figure 1: Schematic representation of the proposed method.
The final objective function as follows:

\[
\min_{U, V} Q(U, V) = \lambda \left( \| Y - X_{CT}^T U \|^2_F + \| Y - X_{CV}^T V \|^2_F \right) \\
+ (1 - \lambda) \left( \| X_{CT}^T U - X_{CT}^T V \|^2_F \right) \\
+ \alpha [tr(U^T X_{CT} L_{CT} X_{CT}^T U)] \\
+ tr(V^T X_{CV} L_{CV} X_{CV}^T V)] \\
+ \beta \left( \| U \|^2_F + \| V \|^2_F \right).
\]

(9)

2.2.2. Optimization Algorithm. Both \( U \) and \( V \) are initialized as zero matrices. We then iteratively update each variable by fixing another variable. By setting the partial derivative of \( Q(U, V) \) with respect to \( U \) and setting it to zero, we have the following:

\[
\frac{\partial (U, V)}{\partial U} = 2\lambda (X_{CT} X_{CT}^T U - X_{CT} Y) + 2(1 - \lambda) (X_{CT} X_{CT}^T Y - X_{CT} X_{CT}^T U) \\
+ 2\alpha X_{CT} L_{CT} X_{CT}^T U + 2\beta U = 0.
\]

We can get the following:

\[
U = (X_{CT} X_{CT}^T + \alpha X_{CT} L_{CT} X_{CT}^T + \beta I)^{-1} [\lambda X_{CT} Y + (1 - \lambda) X_{CT} X_{CT}^T V].
\]

(11)

Similarly, by fixing \( U \) and updating \( V \), we can obtain the following:

\[
V = (X_{CV} X_{CV}^T + \alpha X_{CV} L_{CV} X_{CV}^T + \beta I)^{-1} [\lambda X_{CV} Y + (1 - \lambda) X_{CV} X_{CV}^T U].
\]

(12)

The procedure of projection matrices learning with auxiliary data is described in Algorithm 1.

2.2.3. Feature Extraction of MCI Subjects. Let \( Z_{CT} = [z_{11}^CT, z_{12}^CT, \ldots, z_{m}^CT] \in \mathbb{R}^{d \times m} \) and \( Z_{CV} = [z_{11}^CV, z_{12}^CV, \ldots, z_{m}^CV] \in \mathbb{R}^{d \times m} \) denote the cortical thickness and cortical volume feature matrices of the \( m \) images of MCI subjects, respectively. The feature representations of MCI subjects in subspace are denoted by \( \text{Fea}_{CT} \in \mathbb{R}^{m \times c} \) and \( \text{Fea}_{CV} \in \mathbb{R}^{m \times c} \), which are computed as follows:

\[
\text{Fea}_{CT} = Z_{CT}^T \times U,
\]

(13)

\[
\text{Fea}_{CV} = Z_{CV}^T \times V.
\]

(14)

To make the projected features of pMCI and sMCI subjects are more discriminative, as well as balance the effectiveness of features from thickness and volume data, we perform self-weighted operation and weighted fusion on the features in subspace to obtain the final features. Finally, the biomarkers for MCI subjects are defined as follows:

\[
\text{Fea} = \eta \ast ([\text{Fea}_{CT} \ast \text{Fea}_{CT}]) + (1 - \eta) \ast ([\text{Fea}_{CV} \ast \text{Fea}_{CV}]),
\]

where \( \eta \) is the weight parameter. \(|\text{Fea}_{CT}| \) represents the absolute values of all elements in matrix \( \text{Fea}_{CT} \).
and the global grading biomarker developed in [15], separately for the three classifiers. The classification performances of our proposed biomarkers were significantly better ($p < 0.05$) than that of global grading biomarker in [15] under decision tree classifier and SVM with linear kernel. There was no significant difference in the classification performance between the two competitive biomarkers using SVM with RBF kernel, although the classification accuracy, sensitivity, specificity, and AUC of the proposed biomarkers were slightly higher. In conclusion, the proposed biomarkers were superior to or at least as good as the global grading biomarker in [15] under different classifiers. The proposed biomarker achieved highest accuracy of 69.37% when using SVM classifier with linear kernel.

As mentioned above, the proposed method could reduce the feature dimensions and extract meaningful biomarkers. To verify its performance on dimensionality reduction, we further compared the proposed method with four commonly used feature reduction methods, i.e., minimum redundancy and maximum relevance (mRMR) [51], $t$-test, principal component analysis (PCA) [52], and ICA. The mRMR method selects features according to the minimum redundancy and maximum relevance criterion based on mutual information. $t$-test is one of the statistical hypothesis testing techniques, which has been successfully used for supervised feature selection in neuroimaging studies [53]. Both PCA and ICA are subspace learning methods. PCA captures most of the variance in the data by linearly transforming correlated features into a smaller number of uncorrelated features. ICA separates data into a set of independent and relevant features. We compared above feature reduction methods with the three aforementioned classifiers. The best number of features for each competing method was found by grid search optimization. As can be seen from Figure 2, the proposed method outperformed other feature reduction methods with all three classifiers. The proposed method improved the classification accuracy on average by 9.21%, 8.38%, 7.97%, and 6.43% compared to mRMR, $t$-test, PCA, and ICA, respectively.

3.2. Comparison with State-of-the-Art Methods. In this subsection, we compared the best classification performance of the proposed biomarkers with that of the feature extraction methods presented in [13, 15] on the same dataset. In [13], the MFN features were extracted, and then, the two-step feature selection mRMR and SVM-based recursive feature elimination (SVM-RFE) [54] were employed to find the optimal MFN feature subset. Finally, the SVM classifier with RBF kernel was used for pMCI and sMCI classification. In [15], grading biomarkers were calculated using elastic net technique, and then, the SVM classifier with linear kernel was used for classification. In order to show the validity of our feature extraction strategy, the original morphological features were also added for comparison using the same feature selection strategy and classifier as literature [13].

Table 3 summarizes the classification results of all competing methods. It is notable from Table 3 that all feature extraction methods outperformed the method of exploiting original morphological features in terms of ACC, SPE, and AUC, which implies that the extraction of effective features.

![Algorithm 1: Projection matrices learning algorithm based on auxiliary data.](image-url)
can improve classification performance. In virtue of subspace learning, our proposed method achieved the highest classification accuracy and sensitivity among all competing methods. Specifically, compared with the methods proposed in [13, 15], our method improved the classification accuracy by 3.76% and 1.94% and improved the sensitivity by 4.76% and 5.35%, respectively. Therefore, it is reasonable to integrate subspace learning into the feature extraction, which can enhance the classification power of the features.

The best parameter combination found by experiments was $\lambda = 0.1$, $\alpha = 0.1$, $\beta = 10$, and $\eta = 0.03$. The numbers of nearest neighbors for cortical thickness and volume data were 11 and 3, respectively. For the classification of pMCI and sMCI, the class indicator $c = 2$.

### 3.3. Effectiveness of Learning Projection Matrices Using AD and NC Information

In this subsection, we examined the effectiveness of learning projection matrices using AD and NC data. For comparison, we learned projection matrices by virtue of pMCI and sMCI data. The same procedure of MCI feature extraction as Section 2.2 was conducted. Three different classifiers, i.e., decision tree classifier, SVM with RBF kernel, and SVM with linear kernel, were used for test in turn. We also conducted 10-fold cross-validation for 20 times to obtain the average results. To be specific, we randomly divided the MCI dataset into 10 subsets and then iteratively left one subset for testing and the remaining 9 subsets for training until each of the 10 subsets was validated. The two projection matrices were learned from the training subsets, and then, all the data of training subsets and testing subsets were projected from original space into the subspace by the two projection matrices. At last, the biomarkers were computed according to Equation (15). All the parameters of the competing methods were optimized in the same range as our proposed method.

Table 4 demonstrates the classification results of learning projection matrices using different data. Compared with pMCI and sMCI data, the projection matrices learned with AD and NC data obtained better classification performance no matter which classifier was used. In particular, compared to learning projection matrices using pMCI and sMCI data, the proposed method obtained significant improvements on the classification accuracy and sensitivity by 4.59% and 7.8% when using SVM classifier with linear kernel, respectively. These results confirmed the efficacy of adopting AD and NC data in the subspace learning in our method. Meanwhile, this also validated that the inclusion of AD and NC information is beneficial for the classification between pMCI and sMCI [14, 15, 17, 19, 21, 22, 55].

### 3.4. Visualization

In this subsection, we illustrated the distributions of MCI samples in original morphological feature space and the projected subspace, respectively, to visually exhibit the distinguishing ability of different features. For the original morphological features, the PCA was applied to converting the original thickness and volume features to

![Figure 2: Comparison of different feature reduction methods.](image)

**Table 3: Comparison of classification results of all competing methods.**

<table>
<thead>
<tr>
<th>Feature</th>
<th>Classifier</th>
<th>ACC (%)</th>
<th>SEN (%)</th>
<th>SPE (%)</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original morphological features</td>
<td>SVM classifier with RBF kernel</td>
<td>62.92</td>
<td>71.23</td>
<td>51.72</td>
<td>0.6508</td>
</tr>
<tr>
<td>MFN in [13]</td>
<td>SVM classifier with RBF kernel</td>
<td>65.61</td>
<td>70.63</td>
<td>58.95</td>
<td>0.6670</td>
</tr>
<tr>
<td>Global grading biomarker in [15]</td>
<td>SVM classifier with linear kernel</td>
<td>67.43</td>
<td>70.04</td>
<td>63.96</td>
<td>0.6981</td>
</tr>
<tr>
<td>Proposed biomarker</td>
<td>SVM classifier with linear kernel</td>
<td>69.37</td>
<td>75.39</td>
<td>61.23</td>
<td>0.6951</td>
</tr>
</tbody>
</table>
a number of uncorrelated features, respectively. Here, we employed the first principal component with the largest amount of variance for each type of morphological features and displayed the sample distribution in the two-dimensional space. In the original feature space (Figure 3(a)), it is clear to see that the distributions of pMCI and sMCI samples overlapped severely and samples in each class were scattered. Thus, the classification performance of the original features was very limited. In contrast, interclass distance of the pMCI and sMCI samples in the subspace is large while the intraclass distance is small (Figure 3(b)). Therefore, the proposed biomarkers derived from morphological features exhibited superiority over their original form; that is, our proposed biomarker extraction method was effective. Moreover, from Figures 3(c) and 3(d), we can see that the differences between pMCI and sMCI along the two dimensions in subspace were significant.

4. Discussion

In this work, we presented a novel biomarker extraction method based on subspace learning for the prediction of MCI-to-AD conversion. The developed biomarkers outperformed the competing biomarkers on the discrimination between pMCI and sMCI subjects. Moreover, the improvement from the developed biomarkers was not limited to a particular classifier but worked equally well for three different classifiers. In a word, this work provided a promising biomarker for the early diagnosis of AD.

4.1. Effectiveness Analysis of the Proposed Method. The good performance of our proposed method can be attributed to three reasons: (1) Effective subspace learning. We have demonstrated that the MCI subjects in original morphological feature space were high-dimensional and severely overlapped with each other. Therefore, subspace learning methods mapped multivariate MRI data of MCI subjects into a common subspace with fewer dimensions, where they were much easier to be distinguished. Figure 3 clearly exhibits the efficacy of the space transformation. (2) The information of AD and NC subjects was employed. Compared with MCI subjects, the distances between intraclass samples are small while interclass samples are large for AD and NC subjects. Thus, it is easier to keep the neighborhood relationship between intraclass samples in subspace learning using AD and NC data. In addition, the utilization of AD and NC subjects instead of MCI subjects during subspace learning can avoid the double-dipping problem [56] in the classification of sMCI and pMCI. Therefore, it is reasonable to learn projection matrices using AD and NC data for MCI data, which was verified by the results in Table 4. (3) The self-weighted operation and weighted fusion were conducted. According to the projection matrices learned from AD and NC data, we mapped the thickness and volume data of MCI subjects into a common subspace. The feature representations of MCI subjects in the subspace, i.e., Fea\textsubscript{CT} and Fea\textsubscript{CV}, were obtained. After that, we conducted the self-weighted operation on Fea\textsubscript{CT} and Fea\textsubscript{CV}, to further amplify the differences between pMCI and sMCI. Although the cortical thickness and volume provided complementary information for the discrimination between pMCI and sMCI, the effect of them on classification is imbalanced; the more discriminative the morphological features are, the larger weights they should possess. Thus, we performed weighted fusion on thickness and volume-based features to obtain the final biomarkers. The results in Section 3 implied the effectiveness of our extracted biomarkers.

4.2. Influence of the Number of Auxiliary Data on Classification Accuracy. To study the influence of the number of auxiliary data on classification accuracy, we firstly used different numbers of auxiliary data to calculate the grading biomarker in [15] and the proposed biomarker, respectively, and then compared the differences of performance between them using SVM classifier with linear kernel. The number of auxiliary data varied from 50 to 250 with an increment of 50. For each specific number, we resampled the AD and NC subjects with the proportion of 1:1 for 10 times and calculated the average classification accuracy to avoid the sampling bias. The same procedure of 10-fold cross-validation and parameter optimization as Section 3 were conducted in the classification. The classification accuracies of two competing biomarkers with respect to different numbers of auxiliary data are illustrated in Figure 4. For comparison, we also plotted the classification accuracies of biomarkers computed by all auxiliary data. As shown in Figure 4, the classification performance of both two methods improves gradually with the increase of the number of auxiliary data, which verify the number of auxiliary data has an impact on classification performance of the biomarkers. In addition, the proposed biomarker outperforms the grading biomarker in [15] with different numbers of auxiliary data, which confirms the effectiveness of our proposed method.

4.3. Limitations. There are several limitations that should be addressed in the future work. Firstly, in our work, the CCA

<table>
<thead>
<tr>
<th>Data of learning projection matrices</th>
<th>Classifier</th>
<th>ACC (%)</th>
<th>p value</th>
<th>SEN (%)</th>
<th>SPE (%)</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>pMCI and sMCI data</td>
<td>Decision tree classifier</td>
<td>60.89</td>
<td>&lt;0.0001</td>
<td>67.22</td>
<td>52.27</td>
<td>0.5318</td>
</tr>
<tr>
<td>AD and NC data</td>
<td>Decision tree classifier</td>
<td>66.35</td>
<td>&lt;0.0001</td>
<td>70.97</td>
<td>60.33</td>
<td>0.6216</td>
</tr>
<tr>
<td>pMCI and sMCI data</td>
<td>SVM classifier with RBF kernel</td>
<td>65.17</td>
<td>&lt;0.0001</td>
<td>73.95</td>
<td>53.46</td>
<td>0.6495</td>
</tr>
<tr>
<td>AD and NC data</td>
<td>SVM classifier with linear kernel</td>
<td>68.37</td>
<td>&lt;0.0001</td>
<td>83.57</td>
<td>48.06</td>
<td>0.6800</td>
</tr>
<tr>
<td>pMCI and sMCI data</td>
<td>SVM classifier with linear kernel</td>
<td>64.78</td>
<td>&lt;0.0001</td>
<td>67.59</td>
<td>60.91</td>
<td>0.6628</td>
</tr>
<tr>
<td>AD and NC data</td>
<td>SVM classifier with linear kernel</td>
<td>69.37</td>
<td>&lt;0.0001</td>
<td>75.39</td>
<td>61.23</td>
<td>0.6951</td>
</tr>
</tbody>
</table>

Table 4: Classification performance of learning projection matrices using different data.
Figure 3: Visualization of all MCI samples in original feature space and subspace. (a) The distributions of pMCI and sMCI samples in original morphological feature space. (b) The distributions of pMCI and sMCI samples in subspace. (c, d) The distribution of subject locations by group along the first and second dimensions, respectively.

Figure 4: Classification accuracy of using different numbers of auxiliary data.

was adopted to maintain the correlation between the thickness features and volume features of the same image. And the graph regularization term was used to preserve the neighborhood relationship of samples in the subspace. However, other subspace learning methods, such as ICA, LDA, and LLE, should be further explored and validated in the biomarker extraction. Secondly, to map the MCI subjects into subspace, we learned the projection matrices only using
the information of AD and NC subjects. It remains to be explored whether the performance can be improved by integrating the information of AD, NC, and MCI subjects during the projection matrices learning process. Thirdly, the proposed method took advantage of the limited morphological features, i.e., thickness and volume. As a matter of fact, different morphological features could reflect abnormal alterations of the brain from different perspectives, so they might provide complementary information for the early recognition of disease. More morphologies such as surface area [57], gyrus height [58], and local gyrification index [59] could be adopted to improve the classification performance.

5. Conclusion

In this paper, we developed the novel biomarkers based on subspace learning and the information integration of AD and NC subjects, which found a good feature representation from high-dimensional MRI data for predicting conversion from MCI to AD. The extracted biomarkers exhibited promising performance on discrimination between pMCI and sMCI, which validated the effectiveness of our proposed method. In addition, experimental results showed that the subspace learning was effective approach for finding satisfactory biomarkers and the information integration of AD and NC subjects was beneficial for the prediction of MCI-to-AD conversion.

Data Availability

The data comes from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu/).

Conflicts of Interest

The authors declare no competing interests.

Acknowledgments

This work was supported in part by the National Key Research and Development Program of China (Grant No. 2019YFA0706200), in part by the National Natural Science Foundation of China (Grant No. 61632014 and No. 61627808), in part by the Key Research and Development Program of Shandong Province (Grant No. 2019GGX101056), and in part by the Natural Science Foundation of Shandong Province (Grant No. ZR2019MG022). 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).

References


