

A Novel Grading Biomarker for the Prediction of Conversion from Mild Cognitive Impairment to Alzheimer's Disease

Tong Tong, Qinquan Gao*, Ricardo Guerrero, Christian Ledig, Liang Chen
Daniel Rueckert and the Alzheimer's Disease Neuroimaging Initiative (ADNI)

Abstract—Objective: Identifying mild cognitive impairment (MCI) subjects who will progress to Alzheimer's disease is not only crucial in clinical practice, but also has a significant potential to enrich clinical trials. The purpose of this study is to develop an effective biomarker for an accurate prediction of MCI-to-AD conversion from magnetic resonance (MR) images.

Methods: We propose a novel grading biomarker for the prediction of MCI-to-AD conversion. First, we comprehensively study the effects of several important factors on the performance in the prediction task including registration accuracy, age correction, feature selection and the selection of training data. Based on the studies of these factors, a grading biomarker is then calculated for each MCI subject using sparse representation techniques. Finally, the grading biomarker is combined with age and cognitive measures to provide a more accurate prediction of MCI-to-AD conversion.

Results: Using the ADNI dataset, the proposed global grading biomarker achieved an area under the receiver operating characteristic curve (AUC) in the range of 79%-81% for the prediction of MCI-to-AD conversion within 3 years in 10-fold cross validations. The classification AUC further increases to 84%-92% when age and cognitive measures are combined with the proposed grading biomarker.

Conclusion: The obtained accuracy of the proposed biomarker benefits from the contributions of different factors: a tradeoff registration level to align images to the template space; the removal of the normal aging effect; selection of discriminative voxels; the calculation of the grading biomarker using AD and normal control groups; the integration of sparse representation technique and the combination of cognitive measures.

Significance: The evaluation on the ADNI dataset shows the efficacy of the proposed biomarker and demonstrates a significant contribution in accurate prediction of MCI-to-AD conversion.

Index Terms—structural MR imaging, biomarker, prediction of MCI conversion, machine learning, Alzheimer's disease

I. INTRODUCTION

Studies have shown that individuals with mild cognitive impairment (MCI) progress to Alzheimer's disease (AD) at a rate of 10% to 15% per year [1]. In the clinical context,

T Tong, R. Guerrero, C. Ledig, L. Chen, and D. Rueckert are with Biomedical Image Analysis Group, Department of Computing, Imperial College London, UK.

*Q. Gao is with Fujian Provincial Key Lab of Medical Instrument & Pharmaceutical Technology, Department of the Internet of Things, Fuzhou University, Fuzhou, China.(email: gqinquan@fzu.edu.cn)

Data used in the preparation of this article were obtained from the ADNI database (www.loni.ucla.edu/ADNI). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: www.loni.ucla.edu/ADNI/Collaboration/ADNI_Authorship_list.pdf

it is particularly interesting to predict whether a subject with MCI will go on to develop AD or not within a given time period (i.e. 3 years) as early diagnosis would allow doctors to treat patients sooner so that potential disease-modifying therapies could be tested and applied. Although no drug or treatment has so far been reported to be able to stop the progress of AD, there are some medications that can delay the onset of some symptoms such as memory loss, confusion, and cognitive problems [2]. In addition, the early diagnosis is also helpful for selecting suitable patients for clinical trials. Numerous studies [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20] use neuroimaging techniques to detect pathology associated with AD and to predict the MCI-to-AD conversion. Among them, structural magnetic resonance imaging (MRI) has been the most extensively used imaging modality in the detection and prediction of AD as it is widely available and offers good diagnostic accuracy with moderate costs. In addition, MRI shows high correlation with the progression from MCI to AD [21]. However, the pathological variations between stable MCI (SMCI) and progressive MCI (PMCI) that can be detected by MRI are subtle. This subtle difference is also accompanied with large inter-subject variability and age-related changes, which makes the MRI-based prediction of MCI-to-AD conversion a challenging task. In a recent comprehensive study [22], ten methods were evaluated for the MRI-based prediction of MCI-to-AD conversion and only four methods could discriminate between SMCI and PMCI more accurately than a random classifier. Therefore, it is essential to develop more advanced methods for the prediction of MCI-to-AD conversion.

In order to reduce the large inter-subject variability, MR images are usually spatially registered to a common space for comparison [14], [20], [23]. Different studies have used different registration techniques to align anatomies at different levels of detail. For example, affine registration was used in [23] while more accurate nonrigid registrations were utilized in other studies [14]. In [22], the registration in SPM5 was compared to the more accurate registration method DARTEL. The use of DARTEL was shown to significantly improve the classification results [22]. Although accurate registration can successfully align anatomical structures in different subjects for comparison, the pathological changes between groups may be partially removed at the same time. At the extreme where no registration is performed, the inter-subject variability and the pathological changes between groups are fully preserved.

However, large inter-subject variability may conceal the more subtle pathological variations, hampering the classification of MCI subjects. At the other extreme where registration is near perfect, inter-subject variability and pathological changes are both removed. Thus, there is a tradeoff between the level of image alignment that removes uninteresting inter-subject variability and the amount of detectable pathological changes, which would still allow the measurement of group differences for classification.

The group differences are not only affected by pathological changes and inter-subject variability, but also influenced by the changes associated with aging. Since the normal aging can result in atrophy on the brain, it can be similar to the atrophies caused by AD. This adds confounding effects in using the disease-specific changes for classification. In previous studies, age has been demonstrated as an important feature in either detecting AD [23] or predicting MCI-to-AD conversion [13] because it is a risk factor for AD. Therefore, it may be helpful to remove the effect of normal aging from MRI data before training the classifiers. For example, in a recent challenge [24], eight out of fifteen teams incorporated aging effects in their algorithms for the classification of AD, where the normal aging effect was eliminated using age-dependent normalization [24] or regression [20], [24]. The removal of effects associated with normal aging has been reported to improve the classification performance significantly [20], [25] or slightly [22] depending on the adopted features and classification methods.

In addition to the effect of registration and normal aging, the use of training data from different groups of subjects can also influence the prediction of MCI-to-AD conversion. Since the population of MCI subjects is highly heterogeneous, previous studies [14], [20], [23], [26], [27], [28] have shown that the inclusion of AD and normal controls (NC) subjects can be beneficial for the classification between SMCI and PMCI. A semi-supervised learning method [20], [27] was used to integrate information from AD and NC subject to augment the prediction of MCI-to-AD conversion. Auxiliary data from AD and NC subjects were used in a domain transfer learning method [28] to aid the classification of MCI subjects. In [14], classifiers learned for separating AD and NC subjects were directly used to distinguish between SMCI and PMCI. In [23], new biomarkers were derived by propagating disease information from AD and NC subjects to MCI subjects.

Another important factor for the prediction of MCI-to-AD conversion is feature selection. The number of available voxel-wise features (1,827,095 after preprocessing) in the whole brain is very large, but many of these features may not be related to pathological change due to AD. Therefore, feature selection is an essential step to select the most informative and relevant features for training effective classifiers. Statistical methods such as t-tests or wrapper-based methods such as the support vector machine recursive feature elimination (SVM-RFE) are commonly used approaches for selecting features. More recently, sparse regression techniques have been utilized to select discriminative features for analysis. For example, LASSO regression that incorporates an L1 penalty to encourage a sparse solution has been applied to identify important voxels in the brain that are related to pathological

changes of AD in [29]. However, the selected voxels are randomly distributed over the whole brain, which is difficult for interpretation. More advanced sparse regression methods such as Elastic Net [30] or tree structured sparse learning [29] were proposed to integrate the spatial structure in the images into the feature selection process so that more interpretable results can be generated. In another recent work [31], group sparse regression method has been utilized to select the multimodal features by considering the intrinsic relatedness between features across different modalities.

In this paper, we will first study the impact of the above four different factors: registration accuracy, age correction, selection of training data and features. Based on the observation of the impact of these factors, we further propose a novel biomarker which we refer to as global grading for the prediction of MCI-to-AD conversion. The proposed grading biomarker benefits from the studies of these factors. To calculate the grading biomarker, MR images are first aligned to the template space using well controlled registration. Then, the normal aging effect is removed from the image intensities by using a linear regression model. After that, a feature selection step is carried out to select voxels that can characterize the atrophy associated with AD. Finally, a global grading value is calculated for each MCI subject by utilizing the selected voxels and the disease information of AD and NC subjects. The main novelties of the proposed MR-based grading biomarker are: (1) a well controlled registration accuracy to align MR images to a template space for comparison; (2) the removal of confounding effects associated with aging; (3) a robust feature selection step which is performed on discriminating NC versus AD, thus eliminating the use of MCI subjects in the feature selection step, and (4) the propagation of disease information from NC and AD subjects to MCI subjects using sparse representation technique. The classification performance of the proposed biomarker was validated on the standardized list [32] of baseline scans from the Alzheimer's Disease Neuroimaging Initiative (ADNI) and compared with those of other state-of-the-art methods [20], [22], [33] on the same datasets. Furthermore, the proposed grading biomarker is combined with cognitive measures to gain further improvements in the prediction of MCI-to-AD conversion.

II. MATERIAL AND METHODS

A. ADNI Dataset

Data used in the preparation of this article were obtained from the ADNI database (adni.loni.ucla.edu). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a \$60 million, 5-year public-private partnership. 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 MCI and early AD. Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and

monitor their effectiveness, as well as lessen the time and cost of clinical trials.

The standardized list of baseline scans from ADNI-1 [32] was utilized in our work to enable future comparisons with other methods. The standardized list consists of 818 subjects that were classified into five different groups as shown in Table I. The AD and NC groups were determined according to the diagnosis at baseline. Subjects in the MCI group were classified as PMCI if the subjects converted to AD during a 3 year follow-up. Subjects were classified as SMCI if the diagnosis was MCI at both baseline and 36 months. Those whose diagnosis were missing at 36 months were grouped as unknow MCI (uMCI). The uMCI subjects were not used in the validation.

B. Image Preprocessing

The framework of the proposed classification method is illustrated in Figure 1, consisting of five steps in the preprocessing stage. T1-weighted Magnetic Resonance (MR) brain images were first preprocessed by the standard ADNI pipeline as described in [34], and skull-stripped using the method proposed in [35]. In order to investigate the effect of registration on the prediction of MCI-to-AD conversion, all images were aligned to the MNI152 template using registration based on B-spline free-form deformation (FFD) [36]. Five different levels of deformation were used: affine, 20mm, 10mm, 5mm, 2.5mm (from coarse to fine registration), generating images in the MNI152 template space with different alignment levels. Except for the affine registration, non-rigid were performed using B-spline FFD with the above varying control point spacings. It should be mentioned that if more accurate registration is applied, the images are better aligned for comparison while at the same time more pathological changes might be removed. Additionally, the approach proposed in [37] was used to normalize the image intensities between the subjects and the template. The basic idea in [37] is to deform the histogram of a given image to match a standard histogram. In our work, we used the histogram of the MNI152 template image as the standard histogram. After registration and intensity normalization, all the images were in the same template space and had the same intensity scale for comparison. Then, the normalized intensities in the images can be extracted as features for classification.

The effects of normal aging were then estimated on the voxel-wise intensities of NC subjects by fitting a linear regression model [25] at each location. Given that there are N healthy subjects and each image contains M brain voxels, the normalized intensity values of the preprocessed MR images of all NC subjects can be represented by a matrix $\mathbf{X}^T = [X_1^T, X_2^T, \dots, X_i^T, \dots, X_M^T] \in R^{N \times M}$, where each row is given as $X_i = [x_{i1}, x_{i2}, \dots, x_{iN}] \in R^{1 \times N}$, $i = 1, 2, \dots, M$. The ages of N healthy subjects are denoted as a vector $A \in R^{1 \times N}$. The effect of age-related atrophy is estimated at each voxel separately by fitting a linear regression model $X_i = w_i A + b_i$. After this model is solved in the least squares sense, the aging effect at the i th voxel of the j th image can be removed as $x_{ij}^{new} = x_{ij} - w_i a_j - b_i$. Figure 2 shows an example of the images after different preprocessed steps.

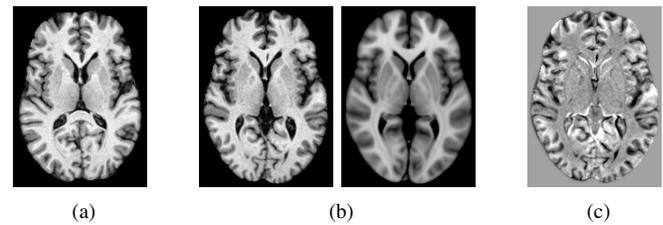


Fig. 2. Images after different preprocessing steps: (a) image after skull-stripping; (b) image after spatially and intensity normalization (left) and the MNI152 template (right); (c) image after the removal of normal aging effect.

C. Feature Selection

After preprocessing, there is still a large amount of voxels available in brain MR images, but not all of them are related to the pathological changes specific to AD. In addition, the number of available images is typically much smaller than the number of voxels, which could result in classifier overfitting. Therefore, a feature selection step is necessary to tackle this problem and also to select informative voxels for classification. Sparse regression techniques have been widely used to select discriminative voxels for AD classification in previous works [29], [30]. The LASSO method using L_1 norm [29] is one of the effective sparse regression methods. However, there are two drawbacks of the original LASSO sparse regression [38]: (a) the structural relationship of neighboring voxels are ignored; (b) the number of selected voxels is limited by the number of images. In order to address these drawbacks, advanced sparse regression methods [29], [30], [31] have been recently proposed for selecting discriminative voxels. For example, manifold learning techniques [31] have been successfully incorporated into the sparse feature selection models to aid the classification. Furthermore, some studies [39], [40] proposed methods of joint sparse feature selection for both disease diagnosis and clinical score prediction. Although these advanced methods [31], [39], [40] provide a good solution for selecting meaningful features, they were applied to select features from hundreds of region-wise features. When feature selection is performed over millions of voxel-wise features, these advanced methods might not be implemented efficiently. In this work, we utilized an Elastic Net (EN) method [30] for feature selection as this method can still be solved efficiently even with an input of millions of features. It is formulated as follows:

$$\hat{\beta} = \min_{\beta} \frac{1}{2} \|y - \mathbf{X}\beta\|_2^2 + \lambda_s \|\beta\|_1 + \lambda_g \|\beta\|_2^2 \quad (1)$$

where \mathbf{X} is a matrix containing vectorized training images and y includes the clinical labels of these images. The EN method adds a L_2 regularization term to the LASSO regression, which results in a grouping effect to select highly correlated variables (i.e. neighboring voxels) and allows us to select a number of voxels greater than the number of images. Equation 1 can be solved efficiently when $\lambda_g \rightarrow \infty$. As shown in [41], a closed-form solution of minimizing Equation 1 can then be obtained for each variable x_i :

TABLE I
DEMOGRAPHIC INFORMATION OF THE DATASET USED IN THIS STUDY. MMSE: MINI-MENTAL STATE EXAMINATION; CDR-SB: CLINICAL DEMENTIA RATING-SUM OF BOXES; RAVLT: REY'S AUDITORY VERBAL LEARNING TEST; FAQ: FUNCTIONAL ACTIVITIES QUESTIONNAIRE; ADAS-COG: ALZHEIMER'S DISEASE ASSESSMENT SCALE-COGNITIVE SUBTEST.

Group	Number	Age	MMSE	CDR-SB	RAVLT	FAQ	ADAS-cog
NC	229	75.9 ± 5.0	29.1 ± 1.0	0.03 ± 0.12	42.9 ± 10.0	0.14 ± 0.60	6.20 ± 2.91
SMCI	129	74.6 ± 7.5	27.6 ± 1.7	1.34 ± 0.65	35.1 ± 10.2	2.03 ± 3.08	9.65 ± 4.11
PMCI	171	74.5 ± 7.0	26.6 ± 1.7	1.88 ± 0.97	27.1 ± 6.2	5.61 ± 5.11	13.20 ± 4.03
uMCI	98	75.2 ± 8.0	27.0 ± 1.9	1.48 ± 0.87	31.3 ± 9.1	3.08 ± 3.73	11.03 ± 4.39
AD	191	75.3 ± 7.5	23.3 ± 2.0	4.31 ± 1.63	22.9 ± 8.1	13.03 ± 6.85	18.55 ± 6.45

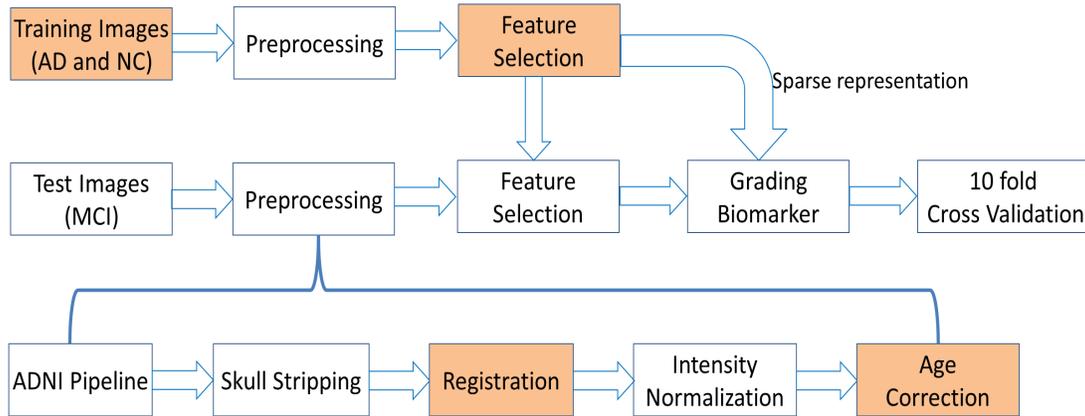


Fig. 1. The framework of the proposed method. The influence of different steps (in orange color) including registration, age correction, feature selection and training groups are studied. Based on the analysis of the impact of these factors, a novel grading biomarker is then proposed for the prediction of MCI-to-AD conversion. The same pre-processing step is applied to all images before calculating the grading biomarker.

$$\hat{\beta}_i = \left(|y^T x_i| - \frac{\lambda_s}{2} \right)_+ \text{sgn}(y^T x_i), i = 1, 2, \dots, M \quad (2)$$

where $(v)_+ = \max(v, 0)$ refers to the positive part. Equation (2) is known as a univariate soft thresholding problem and can be solved very efficiently since $y^T x_i$ is the univariate regression coefficient. Most existing approaches perform the feature selection by just using the data from MCI subjects. However, in a recent study [20], it was shown that the feature selection using data from AD and NC subjects could significantly improve the prediction accuracy of MCI-to-AD conversion. Here, we applied the EN method on images of AD and NC subjects for determining discriminative voxels. Another benefit from using AD and NC subjects for feature selection is that we can avoid the double-dipping problem. The approach that uses the same dataset for selection and selective analysis is referred to as double dipping [42]. For example, if the test data are involved in both the feature selection and the subsequent classifier validation, it will result in the double-dipping problem. Classification methods should avoid double dipping in order to perform a fair validation. In our work, only AD and NC subjects are used for feature selection, avoiding the use of the information from MCI subjects. Thus, our approach does not have this problem. In addition, a re-sampling scheme [30] was employed to reduce sampling bias and to enable a robust feature selection. This scheme selects discriminative voxels by repeatedly fitting the EN regression model on a random subset of the training images while keeping track of voxels that are consistently selected in the sparse

regression model. Figure 3 shows three orthogonal views of the probabilistic mask obtained from the EN algorithm using non-rigid registration with a 10mm control point spacing.

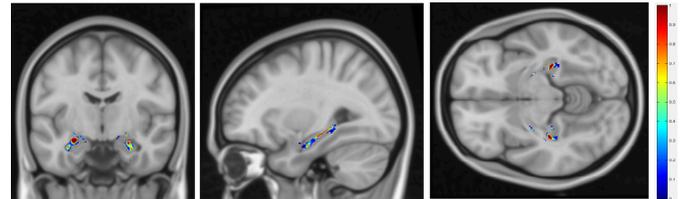


Fig. 3. Orthogonal views of the probability mask for the selection of discriminative voxels.

D. Calculation of Global Grading Biomarker

One aim in this work is to develop a biomarker for the prediction of MCI-to-AD conversion by utilizing the information from AD and NC subjects. As shown in many studies [14], [20], [23], [28], the disease information of AD and NC subjects are helpful for separating PMCI and SMCI. The hypothesis justifying this is that the subjects with SMCI are more NC-like while subjects who go on to develop dementia are more AD-like. The classification of AD vs NC is a simplified version of the classification between PMCI and SMCI. Since the population of MCI subjects is highly heterogeneous, training classifier for AD vs NC might be more effective than training classifier for PMCI vs SMCI. Here, we develop a novel biomarker by propagating disease labels of NC and AD subjects to MCI subjects. A global

grading value is calculated for each MCI subject and used as biomarker for classification. Given the NC and AD subjects as the training population, the relationship between the training population and each MCI subject needs to be investigated so that the disease information of the training population can be propagated to MCI subjects. In [23], the relationship between each MCI subject and the training population is modeled by a weighting function. Here, we model this relationship using a sparse representation method, which has been demonstrated to be superior to the weighting function in image segmentation [43], [44]. Each MCI subject is assumed to lie in the space of the training population, which means that it can be represented by a linear combination of NC and AD subjects in the training population. In addition, a sparse constraint is added to the representation. In order to seek a sparse representation of each MCI subject, we also use the Elastic Net technique. After feature selection, K discriminative voxels are determined and K intensity values are extracted from each image. Given $\mathbf{X}^{ADNC} \in R^{K \times N}$ that contains the intensity values of N training images and $X^{MCI} \in R^{K \times 1}$ that contains the intensity values of a single MCI image, the sparse representation of this MCI subject can be obtained by minimizing the following cost function:

$$\hat{\alpha} = \min_{\alpha} \frac{1}{2} \|\mathbf{X}^{MCI} - \mathbf{X}^{ADNC} \alpha\|_2^2 + \lambda_1 \|\alpha\|_1 + \lambda_2 \|\alpha\|_2^2 \quad (3)$$

Here $\hat{\alpha}$ are the coding coefficients. Most of the coefficients in $\hat{\alpha}$ are zero due to the sparsity constraint. If the coefficient in $\hat{\alpha}$ is not zero, it indicates that the corresponding training image has been selected to propagate its clinical label information to the target MCI subject. By adding L_2 norm in Equation (3), a grouping effect can be obtained over the sparse coding coefficients. Qualitatively speaking, an algorithm exhibits the grouping effect if the coding coefficients of a group of highly correlated subjects tend to be equal. For example, there are two training subjects with very similar or identical intensity patterns. If we just use L_1 norm, it will select one of them while eliminating the other one. However, in calculating the grading biomarker, both subjects are similar to the target subject and should be used to propagate their disease information to the target subject. After adding L_2 norm as in equation (3), both subjects can be selected in calculating the grading biomarker. After the sparse solution is obtained, the scoring of the target MCI subject is based on the coding coefficients $\hat{\alpha}$ and the clinical status of the selected training population. The clinical status of a training image is denoted as s_j . If the training image is a NC subject, s_j is set to 1; otherwise, s_j is set to -1 for AD subjects. A global grading value of the target MCI subject is then calculated by:

$$g^{MCI} = \frac{\sum_{j=1}^N \hat{\alpha}(j) s_j}{\sum_{j=1}^N \hat{\alpha}(j)} \quad (4)$$

where N is the number of the training images in \mathbf{X}^{ADNC} and $\hat{\alpha}(j)$ is the coding coefficient corresponding to the training image $X^{ADNC}(j)$. If g^{MCI} is close to -1 , it indicates that this MCI subject is more characteristic of PMCI than SMCI and is likely to convert to AD within the given time period.

If g^{MCI} is close to 1, it indicates that the MCI subject will possibly not convert to AD and remain stable within the given time period.

E. Implementation details

For training classifiers with single type of features (i.e. intensity features or grading biomarkers), support vector machine (SVM) with a linear kernel was used as SVM is one of the most widely used classifiers. The implementation of SVM was performed using liblinear in matlab (<http://www.csie.ntu.edu.tw/~cjlin/liblinear>). The liblinear was used since it can be implemented efficiently for large-scale linear classification. It is a large-scale classification problem when millions of voxel-wise intensity features are used for the input of classifier training. The linear SVM only has one soft margin parameter C , which was optimized in the range of 2^i , $i = \{-5, -4, \dots, 5\}$ via cross validation. All features were normalized to have zero mean and unit variance before the classification. For calculating the grading biomarker, the parameters λ_1 and λ_2 in Equation 3 were determined on the logarithmic scale 10^i , $i = \{-3, -2, \dots, 3\}$. The best setting found is $\lambda_1 = 0.1$ and $\lambda_2 = 100$, which was used for the final calculation of grading features. The sparse coding process of EN was carried out using the SPAMS toolbox (<http://spams-devel.gforge.inria.fr/index.htm>) [45]. For the evaluation of classification performance, 10-fold cross validation was carried out. The reported results in terms of accuracy (ACC), sensitivity (SEN), specificity (SPE) and area under curve (AUC) are averages over 100 runs. To assess the statistical significance of different results, Student's t-tests were performed using the accuracies of the 100 runs. In addition, in order to allow for a fair comparison with the best result reported in [20], we combined the proposed global grading biomarker with age and cognitive measures for classification. The random forest classifier (<http://code.google.com/p/randomforest-matlab>) was then trained using the combined biomarkers as in [20]. The random forest classifier was utilized because it was shown to yield better performance than SVM for combining different types of biomarkers [20]. Therefore, when the input was an individual type of features, SVM was used for classifier training while random forest was only used in combining of different types of features.

III. RESULTS

A. The impact of registration and age correction

Experiments were conducted to investigate the effect of registration and age correction on the prediction of MCI-to-AD conversion. Since AD and NC subjects were used as training images in the following feature selection step, the impact of registration and age correction were also analysed for the classification of AD vs NC. As can be seen from Figures 4 and 5, the SVM classifiers achieve the best accuracies using registration with a control point spacing between 5mm and 10mm in both classification scenarios, which can be a good choice for the classification of AD using MR intensities. In addition, Figure 4 shows that the removal of the age-related effects is significantly helpful for the classification

of AD vs NC at all registration levels. However, only slight improvements were obtained on the classification of PMCI vs SMCI after aging correction as shown in Figure 5. This is due to the fact that the normal aging effect was estimated on the NC subjects. The estimated effect is biased to the NC group. Thus, the classification of AD vs NC can benefit more from the removal of aging effect than the classification of PMCI vs SMCI.

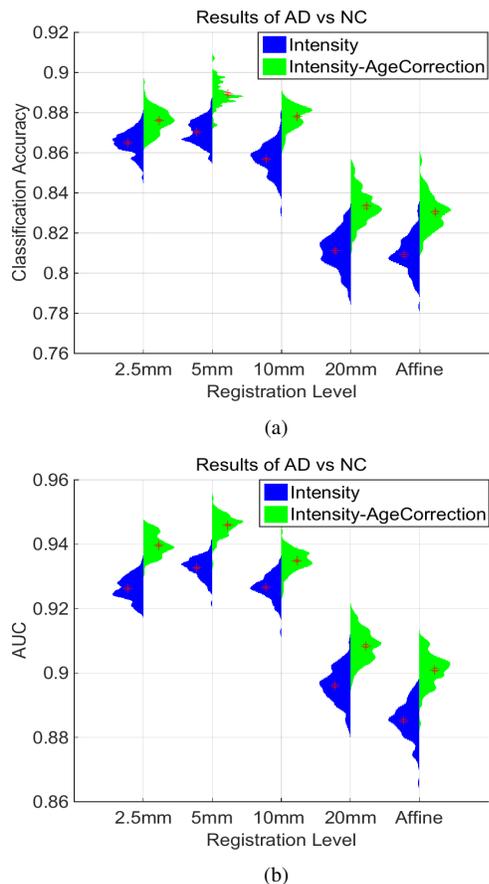


Fig. 4. Results of AD vs NC using different registration levels: (a) classification accuracy; (b) AUC. All the evaluations used normalized intensities as features. The beanplots show the classification results of 100 runs of 10-fold cross validation. SVM was used for training classifiers.

To visually check the impact of registration, we compared the group difference between AD and NC at the same location in the MNI152 template space using t-tests. The p-values were adjusted using the positive false discovery rate method as described in [46]. The adjusted p-value maps at different registration levels are shown in Figure 6. Visually, statistical differences between AD and NC in terms of p-values become significant but also gradually disappear in the temporal lobes as more accurate registration was applied. Specifically, if only affine registration is used, all the p-values are in the red or yellow region of the color bar as shown in Figure 6, which could result from both the inter-subject variability and the pathological changes. However, the inter-subject variability at this registration level may contribute much more to the statistical differences than the pathological changes as the anatomies are not aligned well for comparison by just using

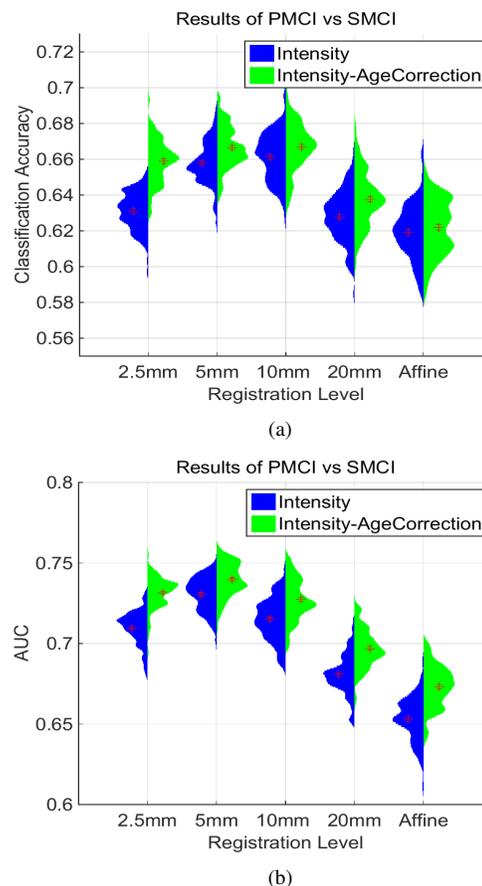


Fig. 5. Results of PMCI vs SMCI using different registration levels: (a) classification accuracy; (b) AUC. All the evaluations used normalized intensities as features. The beanplots show the classification result of 100 runs of 10-fold cross validation. SVM was used for training classifiers.

affine registration. In addition, we observed that the highest discriminative power in the ventricle region was detected using non-rigid registration with a 20mm control point spacing while the most significant differences in the hippocampus was detected using registration with a control point spacing between 5mm and 10mm. Although less discriminative voxels were found using registration with a 2.5mm control point spacing than that using registration with a 5mm or 10mm control point spacing, the remaining discriminative voxels at this registration level still allow the measurement of group differences for classification. This is confirmed by our classification results as shown in Figures 4 and 5. Overall, the registration with a control point spacing between 5mm and 10mm provides a tradeoff between removing the inter-subject variability and keeping the detectable pathological difference for classification. In addition, it should be mentioned that most of the features selected using Equation (1) are in the low p-value region as shown in Figure 6. However, there are also some differences between the two spatial distributions. This is due to the fact that the p-value maps in Figure 6 were calculated using t-tests while the feature selection was performed using more advanced sparse regression method.

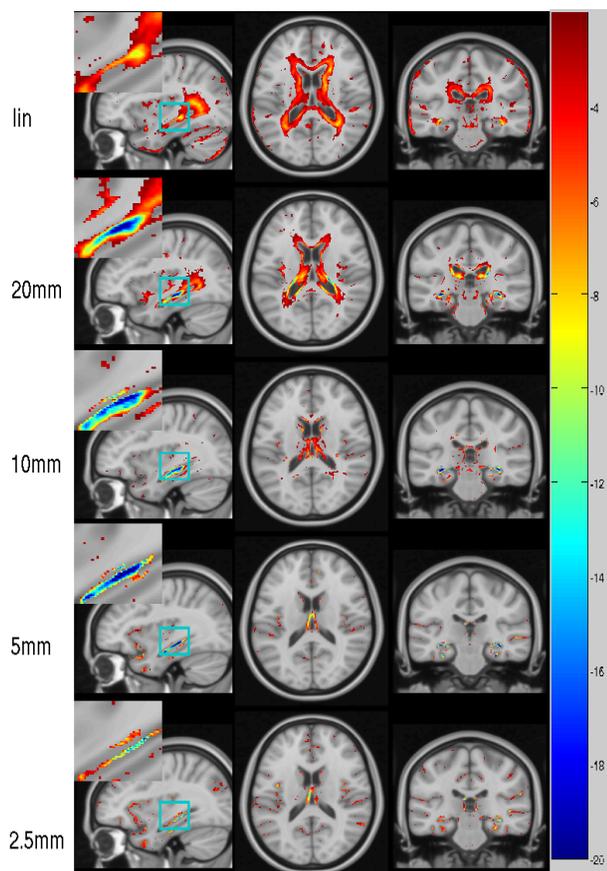


Fig. 6. Visualization of the impact of registration. The presented results were obtained on 191 AD vs 229 NC using t-tests. The normalized intensities after age correction were used for comparison. Only the p-values lower than 10^{-2} are displayed. The showed overlay values are the log base 10 of p-values. The hippocampus area is enlarged for visualization. The lower p values (i.e as shown in blue) represent more significant differences between AD patients and healthy subjects, indicating possible disease changes due to AD in that region (i.e the blue region).

B. The importance of feature selection and training data selection

The performance of the EN feature selection was further evaluated for the classification of PMCI vs SMCI. We also performed feature selection on AD and NC subjects. The intuition behind this is that the pathological changes from NC to AD should cover the pathological changes from SMCI to PMCI. Thus, we assumed that the features selected from discriminating NC vs AD would be useful for the classification between PMCI and SMCI. As shown in Table II, the improvement of using the EN feature selection was the most significant at the registration level with a control spacing of 10mm, but the feature selection is not consistently helpful at all registration levels. In addition, the feature selection performed over AD and NC subjects are more effective than that performed over MCI subjects.

Moreover, the influence of using different data for training classifiers was also studied. We performed both the feature selection and the classifier training on AD vs NC and then applied the learned classifier for discriminating PMCI vs SMCI. The results are shown in Table II. It is interesting to

TABLE II

CLASSIFICATION RESULTS OF PMCI VS SMCI. INTENSITIES WERE USED AS FEATURES. SVM WAS USED FOR TRAINING CLASSIFIERS. THE METHOD MCI_MCI REPRESENTS THAT THE CLASSIFIERS WERE TRAINED ON THE MCI SUBJECTS WITHOUT FEATURE SELECTION. THE METHOD MCIEN_MCI MEANS THAT BOTH THE FEATURE SELECTION USING EN AND CLASSIFIER TRAINING USING SVM WERE PERFORMED ON MCI SUBJECTS. THE METHOD ADNCEN_MCI REPRESENTS THAT THE FEATURE SELECTION WAS PERFORMED ON AD VS NC WHILE THE CLASSIFIER TRAINING WAS CARRIED OUT ON MCI SUBJECTS. ADNCEN_ADNC REPRESENTS THAT BOTH THE FEATURE SELECTION AND CLASSIFIER TRAINING WERE PERFORMED ON AD VS NC.

Method	Registraion level	Affine	20mm	10mm	5mm	2.5mm
MCI_MCI	ACC (%)	63.1	63.7	66.5	66.7	65.9
	AUC (%)	68.3	69.2	72.8	74.0	73.2
MCIEN_MCI	ACC (%)	62.0	63.8	69.1	66.9	64.8
	AUC (%)	67.1	69.0	75.8	73.1	72.7
ADNCEN_MCI	ACC (%)	63.4	65.2	69.2	67.9	66.0
	AUC (%)	69.0	69.7	76.5	75.3	71.6
ADNCEN_ADNC	ACC (%)	66.7	70.0	73.2	72.4	71.7
	AUC (%)	70.8	74.2	78.2	76.5	76.3

note that the use of AD and NC subjects in training classifier can significantly improve the classification accuracy and the improvement is consistently at all registration levels.

C. Grading biomarker

Based on the selected voxels using EN, the disease information of AD and NC subjects were then propagated to MCI subjects to calculate the global grading biomarkers as described in section II-D. In this way, we can avoid training classifiers on AD vs NC but still use the disease information from AD and NC subjects to aid the classification of PMCI vs SMCI due to the propagation process. The proposed global grading biomarker was also evaluated using different levels of registration. In addition, the hippocampus grading biomarker proposed in [23] was calculated for comparison. In order to calculate the hippocampus grading biomarker at different registration levels, several steps were carried out: (1) the hippocampus was segmented in native space using the method described in [47]. In this method, 30 atlases were transformed to a target image space using non-rigid registration [36]. Then, atlas label maps were transformed using the obtained transformations. Finally, a label fusion step with an expectation-maximization (EM) refinement was conducted to obtain a consensus segmentation. The detailed parameter settings of this approach can be found in [47]. (2) the segmented hippocampus was transformed to the template space using different levels of registration; (3) the clinical labels were propagated from NC and AD subjects to MCI subjects at voxel level using patches [47]; (4) average grading value within the hippocampus, which we term the hippocampus grading, was calculated for each MCI subject. Figure 7 shows that the classification results using the proposed global grading are more accurate than those using the hippocampus grading. The improvement is hypothesized to be due to the feature selection using EN and the propagation of disease information at image level.

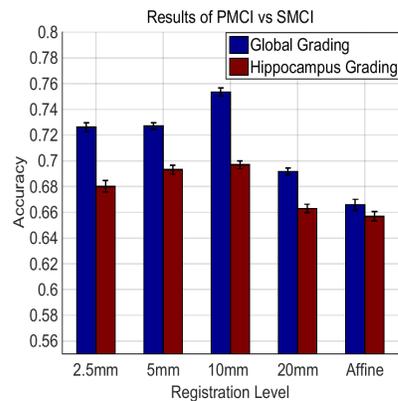


Fig. 7. Comparison of the classification results using the global grading biomarker and the hippocampus (local) grading biomarker at different registration levels. 100 runs of 10-fold cross validation were carried out. SVM was used for training classifiers.

TABLE III

COMPARISON OF CLASSIFICATION RESULTS OF PMCI VS SMCI WITH THE ADNI SUBSET USED BY [22]. BACC REPRESENTS BALANCED ACCURACY, WHICH IS CALCULATED AS THE AVERAGE OF SENSITIVE AND SPECIFICITY.

Method	Classifier	SEN	SPE	AUC	BACC
Voxel-COMPARE in [22]	SVM	62%	67%	-	65%
Voxel-STAND in [22]	SVM	57%	78%	-	67%
Hippo-Volume in [22]	Parzen	62%	69%	-	66%
Thickness-Direct in [22]	SVM	32%	91%	-	62%
All biomarkers in [33]	LDA	69%	54%	-	62%
MRI biomarker in [20]	SVM	64%	72%	75%	68%
Aggrerate biomarker in [20]	RF	40%	94%	81%	67%
Proposed global grading biomarker	SVM	86.5%	67.2%	79.0%	76.9%
Proposed combined biomarkers	RF	85.2%	71.8%	83.8%	78.5%

D. Comparison with state-of-the-art methods

In this study, we evaluated the proposed method with the same data as used by [20] and [22] respectively. The data from these studies was released and are also subsets of the standardized set [32]. The same cross validations as in [20] and [22] were used for fair comparisons. The global grading biomarker was calculated using registration with a control point spacing of 10mm. In addition, the global grading biomarker was combined with age and cognitive measures for classification (including MMSE, CDR-SB, RAVLT, FAQ and ADAS-cog as shown in Table I), resulting in what we call the combined biomarkers. This was performed as it has been previously demonstrated to aid the classification of MCI subjects [20]. In order to ensure a fair comparison with [20], the random forest classifier [20] was used to integrate age and cognitive measures for classification. Different measures of the same subject were concatenated into a feature vector for training. Tables III and IV show that our proposed method achieves competitive performance with these state-of-the-art methods. Based on the dataset used in [22], only four out of ten methods can predict conversion slightly more accurately than a random classifier as shown in Table III but none of them obtained significantly better results. The obtained accuracy of our proposed method is 76.1% with an AUC of 83.8%, which is much higher than that of a random classifier. To the best of our knowledge, the study in [20] reported the highest classification performance on the prediction of MCI-to-AD conversion to

TABLE IV

COMPARISON OF CLASSIFICATION RESULTS OF PMCI VS SMCI WITH THE ADNI SUBSET USED BY [20].

Method	Classifier	ACC	SEN	SPE	AUC
MRI biomarker in [20]	SVM	74.7%	88.9%	51.6%	76.6%
Proposed global grading biomarker	SVM	78.9%	76.0%	82.9%	81.3%
Aggrerate biomarker in [20]	RF	81.7%	86.7%	73.6%	90.2%
Proposed combined biomarkers	RF	84.1%	88.7%	76.5%	91.7%

date. In addition, to enable future comparisons with other methods, we have provided the evaluation script (including image list, features and cross-validation source codes available at <http://scholar.harvard.edu/ttong/software/novel-grading>) to reproduce the best result reported in this paper.

IV. DISCUSSION

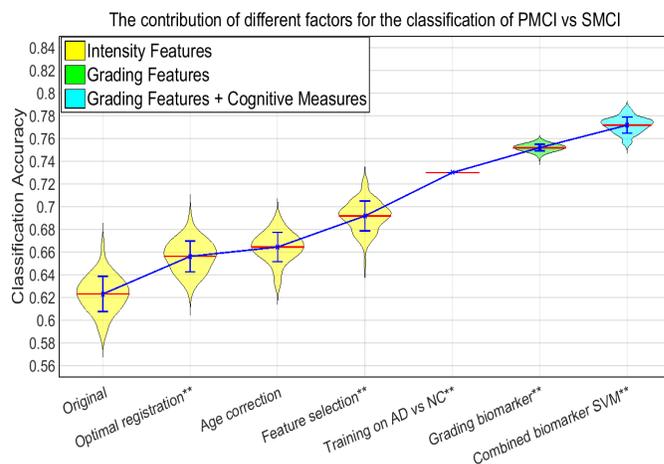


Fig. 8. The classification results of PMCI vs SMCI by adding different factors one by one. SVM was used for training classifiers. ** means that the result is significantly improved by adding this factor. Original means that the results were obtained by using normalized intensities after affine registration. There is only one accuracy for the fifth column because the classifier was trained on discriminating AD vs NC and then directly applied on the classification of PMCI vs SMCI. Thus, there is no cross validation in this scenario. For the sixth column, cross validations were applied to all MCI subjects using their grading biomarkers as input features while the combined biomarkers were used in the seventh column.

In this study, we have developed a novel grading biomarker for the prediction of MCI-to-AD conversion. First, we investigated the impact of several important factors on the prediction performance including registration accuracy, age correction, feature selection and the selection of training data. By adding or optimizing these preprocessing steps, the classification accuracy between SMCI and PMCI can be gradually improved from 62% to 73% as shown in Figure 8. Except for the age correction, all the other three factors significantly contribute to the improvement of the prediction accuracy. Based the analysis of these factors, we further proposed a novel grading biomarker. Using the grading biomarker, the prediction accuracy further increased to 75%. The proposed grading biomarker benefits not only from the studies of the above pre-processing steps but also from the use of the sparse representation technique. Due to the sparsity constraint, the sparse representation technique can only allow a few similar training subjects from AD and

TABLE V
RESULTS OF DIFFERENT CLASSIFIERS USING THE COMBINED BIOMARKERS.

Method	ACC	SEN	SPE	AUC
Combined biomarker SVM	77.2%	75.0%	80.1%	84.9%
Combined biomarker RF	80.7%	86.7%	72.6%	87.0%

NC to propagate their disease information to the MCI subjects for calculating the grading features, thus eliminating some dissimilar subjects in the feature calculation. The obtained accuracy using age and cognitive measures is 76%, which is slight higher than that of the grading biomarker. However, the prediction accuracy was further boosted to 81% with an AUC of 87% by combining the proposed grading biomarker with age and cognitive measures, indicating that there are complementary information between these different measures.

One contribution of our work is the comprehensive study of the impact of different factors: (1) We demonstrated that non-rigid registration with a control spacing point between 5mm and 10mm is a good choice for the classification of AD using MR intensities as this registration level can provide a tradeoff between the removal of inter-subject variability and the amount of detectable pathological changes. However, as shown in Figure 5, the pathological changes in the ventricle region was best detected using registration with a 20mm control point spacing while the highest discriminative power in the hippocampus area was obtained after applying registration with a control spacing point between 5mm and 10mm. Therefore, it may be helpful to apply a multi-level feature selection to extract the pathological changes at different registration levels so that the best detectable pathological changes at different regions can be combined for classification. We will investigate this in our future work. (2) Since normal aging has similar atrophy effects on certain regions as AD [20], [25], [48], it would cause a confounding effect on using the disease-specific changes for classification. Thus, a linear regression model [25], [49] was used to remove the confounding effect of normal aging. Although the improvement by adding age correction is not significant, it is consistently helpful for all the classification experiments as shown in Figures 4 and 5. In addition, it should be noted that the linear correction of age is an approximation of the normal aging process since the trajectory of normal aging can be assumed to be nonlinear as shown in previous studies [21], [50]. A more specialized nonlinear regression model would be more effective in removing the confounding effects of normal aging. Although the normal aging effect was removed from the calculation of disease-specific MR biomarkers, age was added as an additional feature in the final classification. This is due to the fact that age is a risk factor for AD. It is reported that older subjects are more likely to develop AD than younger subjects. This means that age is an important predictor for classification of AD. Thus, we added age as an additional feature in the final classification. (3) Moreover, we observed that the effectiveness of the feature selection is related to the accuracy of the registration as demonstrated in Table II. When affine registration is applied, the anatomies are not aligned well and thus the feature selection is not effective at this registration

level as it does not compare like with like. When very accurate registration is applied (i.e. non-rigid registration with a 2.5mm control point spacing), the feature selection is also not effective as less noisy features present due to the removal of inter-subject variability. Feature selection is more effective at the registration level between 5mm and 10mm since not only the pathological changes can be detected at this registration level but also the amount of noisy voxels (possible due to remaining inter-subject variability) is still large as shown in Figure 5. (4) In addition, when the classifier is trained on AD vs NC, the classification performance between PMCI and SMCI can be significantly improved, which is in accordance with results presented in previous studies [14], [51], [52]. This finding is encouraging because it is easier to obtain labeled training data from AD and NC subjects than from MCI subjects (MCI subjects need to be tracked for years to establish their training labels while the training labels of AD and NC subjects can be determined at baseline).

Another contribution in this work is the proposal of the global grading biomarker. In contrast to recent studies on the ADNI data that rely on combinations of complex features [7], [33], [53], [54], [55] or integrations of multiple modalities [5], [56], [14], the proposed biomarker is a single MRI-based interpretable feature. When the grading value is close to -1 , it indicates that the subject is more characteristic of PMCI than SMCI and has a high possibility to convert to AD within 3 years while the grading value 1 means that this MCI subject will possibly remain stable within this period. Although the proposed grading biomarker is calculated using normalized intensities, it can also be calculated using other features such as the grey matter density maps [22] or cortical thickness [13]. However, the factors including registration, age correction, feature selection and training groups may also have significant influence on these features and need to be studied in order to calculate an effective grading biomarker for the prediction of MCI-to-AD conversion.

The prediction accuracy for conversion from MCI to AD varies in the range of 56%-82% in recent studies as shown in Table VI. The variations in the reported results can be caused by several factors such as different machine learning methods, different biomarkers, different subsets from ADNI, different cross validations. Another important factor, which can also significantly affect the prediction results, is how the MCI subjects are separated into SMCI and PMCI. For example, as shown in Table VI, the proposed combined biomarkers achieved significantly different results when different definitions of SMCI and PMCI were adopted. Based on the definition as described in section II-A, the proposed biomarkers achieved a classification AUC of 87%. When we applied the same validation on the subjects using the definition of SMCI and PMCI as described in [20], the classification AUC increased to 92%. The improvement is caused by the strict definition of SMCI as in [20], which removed those SMCI subjects who converted to AD after 36 months. In spite of the large variation in the reported accuracies, it is encouraging to note that there is an increasing trend in the prediction accuracy with the recent progress on this challenging task, which will finally lead to a prediction rate suitable for clinical use.

TABLE VI
RESULTS OF PREVIOUS STUDIES FOR PREDICTING MCI-TO-AD CONVERSION ON THE ADNI DATASET. RF WAS APPLIED TO THE COMBINED BIOMARKER FOR TRAINING CLASSIFIERS. ACC: ACCURACY; AUC: AREA UNDER THE RECEIVER OPERATING CHARACTERISTIC CURVE.

Study	Data	Number of Subjects	Conversion time	ACC	AUC
[4]	MRI	76 PMCI, 134 SMCI	0-18 months	64%	-
[57]	MRI	72 PMCI, 50 SMCI	0-24 months	73%	-
[3]	MRI	27 PMCI, 76 SMCI	0-36 months	81%	77%
[22]	MRI	76 PMCI, 134 SMCI	0-18 months	67%	-
[33]	MRI	167 PMCI, 238 SMCI	0-48 months	56%-68%	-
[58]	MRI, CSF	69 PMCI, 170 SMCI	0-36 months	62%	73%
[59]	MRI	154 PMCI, 215 SMCI	0-36 months	72%	-
[9]	MRI	62 PMCI, 256 SMCI	0-18 months	66%	65%
[7]	MRI	72 PMCI, 131 SMCI	0-18 months	71%	-
[23]	MRI, age	167 PMCI, 238 SMCI	0-48 months	74%	-
[60]	MRI, PET cognitive scores	38 PMCI, 50 SMCI	0-24 months	78%	77%
[10]	MRI cognitive scores	153 PMCI, 182 SMCI	0-36 months	65%	-
[55]	MRI	89 PMCI, 111 SMCI	0-36 months	75%	84%
[14]	MRI, PET APOE	47 PMCI, 96 SMCI	0-36 months	74%	80%
[19]	MRI	167 PMCI, 238 SMCI	0-48 months	72%	76%
[20]	MRI, age cognitive scores	164 PMCI, 100 SMCI	0-36 months	82%	90%
Combined biomarkers (this study)	MRI, age cognitive scores	164 PMCI, 100 SMCI	0-36 months	84%	92%
		171 PMCI, 129 SMCI	0-36 months	81%	87%

V. CONCLUSION

In this work, we first investigated the influence of four factors on the prediction of MCI-to-AD conversion: registration accuracy, age correction, feature selection and the selection of training data. Furthermore, we have proposed a new biomarker based on global grading for the prediction of MCI-to-AD conversion. The evaluation on the ADNI dataset shows promising results and demonstrates the efficacy of the proposed biomarker. As indicated by the experimental results, the improvement of the classification performance on the MCI-to-AD conversion prediction benefits from the contribution of several factors: (a) the optimal registration level to align images to the template space; (b) the removal of the age-related effect; (c) the feature selection step using EN; (d) the calculation of the grading biomarker using AD and NC groups; (e) the use of the sparse representation technique; (f) the combination of cognitive measures.

ACKNOWLEDGMENTS

This work was partially funded under the 7th Framework Programme by the European Commission (<http://cordis.europa.eu>; EU-Grant-611005-PredictND) and the Provincial Natural Science Foundation of Fujian, China (Grant No. 2016J05157). The ADNI Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI; Principal Investigator: Michael Weiner; NIH grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering (NIBIB), and through generous contributions from the following: Pfizer Inc., Wyeth Research, Bristol-Myers Squibb, Eli Lilly and Company, GlaxoSmithKline, Merck & Co. Inc., AstraZeneca AB, Novartis Pharmaceuticals Corporation, Alzheimer's Association, Eisai Global Clinical Development, Elan Corporation plc, Forest Laboratories, and the Institute for

the Study of Aging, with participation from the U.S. Food and Drug Administration. Industry partnerships are coordinated through the Foundation for the National Institutes of Health. The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory of Neuroimaging at the University of California, Los Angeles.

REFERENCES

- [1] M. Grundman *et al.*, "Mild cognitive impairment can be distinguished from Alzheimer disease and normal aging for clinical trials," *Archives of neurology*, vol. 61, no. 1, pp. 59–66, 2004.
- [2] K. G. Yiannopoulou and S. G. Papageorgiou, "Current and future treatments for Alzheimer's disease," *Therapeutic advances in neurological disorders*, vol. 6, no. 1, pp. 19–33, 2013.
- [3] C. Misra *et al.*, "Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: results from ADNI," *NeuroImage*, vol. 44, no. 4, pp. 1415–1422, 2009.
- [4] M. Chupin *et al.*, "Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI," *Hippocampus*, vol. 19, no. 6, pp. 579–587, 2009.
- [5] D. Zhang *et al.*, "Multimodal classification of Alzheimer's disease and mild cognitive impairment," *NeuroImage*, vol. 55, no. 3, pp. 856–867, 2011.
- [6] C.-Y. Wee *et al.*, "Enriched white matter connectivity networks for accurate identification of MCI patients," *NeuroImage*, vol. 54, no. 3, pp. 1812–1822, 2011.
- [7] Y. Cho *et al.*, "Individual subject classification for Alzheimer's disease based on incremental learning using a spatial frequency representation of cortical thickness data," *NeuroImage*, vol. 59, no. 3, pp. 2217–2230, 2012.
- [8] J. Ye *et al.*, "Sparse learning and stability selection for predicting MCI to AD conversion using baseline ADNI data," *BMC neurology*, vol. 12, no. 1, p. 46, 2012.
- [9] E. Westman *et al.*, "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.

- [10] R. Casanova *et al.*, "Alzheimer's disease risk assessment using large-scale machine learning methods," *PLoS one*, vol. 8, no. 11, p. e77949, 2013.
- [11] C. Gaser *et al.*, "BrainAGE in mild cognitive impaired patients: predicting the conversion to Alzheimer's disease," *PLoS one*, vol. 8, no. 6, p. e67346, 2013.
- [12] K. R. Gray *et al.*, "Random forest-based similarity measures for multimodal classification of alzheimer's disease," *NeuroImage*, vol. 65, pp. 167–175, 2013.
- [13] S. F. Eskildsen *et al.*, "Prediction of Alzheimer's disease in subjects with mild cognitive impairment from the ADNI cohort using patterns of cortical thinning," *NeuroImage*, vol. 65, pp. 511–521, 2013.
- [14] J. Young *et al.*, "Accurate multimodal probabilistic prediction of conversion to Alzheimer's disease in patients with mild cognitive impairment," *NeuroImage: Clinical*, vol. 2, pp. 735–745, 2013.
- [15] G. Yu *et al.*, "Multi-task linear programming discriminant analysis for the identification of progressive MCI individuals," *Plos one*, vol. 9, no. 5, p. e96458, 2014.
- [16] T. Tong *et al.*, "Multiple instance learning for classification of dementia in brain MRI," *Medical Image Analysis*, vol. 18, no. 5, pp. 808–818, 2014.
- [17] F. Li *et al.*, "A robust deep model for improved classification of AD/MCI patients," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 5, pp. 1610–1616, 2015.
- [18] C.-Y. Wee *et al.*, "Sparse temporally dynamic resting-state functional connectivity networks for early MCI identification," *Brain Imaging and Behavior*, pp. 1–15, 2015.
- [19] S. Eskildsen *et al.*, "Structural imaging biomarkers of Alzheimer's disease: predicting disease progression," *Neurobiology of Aging*, vol. 36, pp. 23–31, 2015.
- [20] E. Moradi *et al.*, "Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects," *NeuroImage*, vol. 104, pp. 398–412, 2015.
- [21] C. R. Jack *et al.*, "Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade," *The Lancet Neurology*, vol. 9, no. 1, pp. 119–128, 2010.
- [22] 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. 766–781, 2011.
- [23] P. Coupé *et al.*, "Scoring by nonlocal image patch estimator for early detection of Alzheimer's disease," *NeuroImage: Clinical*, vol. 1, no. 1, pp. 141–152, 2012.
- [24] E. E. Bron *et al.*, "Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: The CADDementia challenge," *NeuroImage*, vol. 111, pp. 562–579, 2015.
- [25] J. Dukart *et al.*, "Age correction in dementia-matching to a healthy brain," *PLoS one*, vol. 6, no. 7, p. e22193, 2011.
- [26] K.-H. Thung *et al.*, "Neurodegenerative disease diagnosis using incomplete multi-modality data via matrix shrinkage and completion," *NeuroImage*, vol. 91, pp. 386–400, 2014.
- [27] R. Filipovych *et al.*, "Semi-supervised pattern classification of medical images: application to mild cognitive impairment," *NeuroImage*, vol. 55, no. 3, pp. 1109–1119, 2011.
- [28] B. Cheng *et al.*, "Domain transfer learning for MCI conversion prediction," in *International Conference on MICCAI*, 2012, pp. 82–90.
- [29] M. Liu *et al.*, "Identifying informative imaging biomarkers via tree structured sparse learning for AD diagnosis," *Neuroinformatics*, vol. 12, no. 3, pp. 381–394, 2014.
- [30] R. Guerrero *et al.*, "Manifold population modeling as a neuro-imaging biomarker: Application to ADNI and ADNI-GO," *NeuroImage*, vol. 94, pp. 275–286, 2014.
- [31] B. Jie *et al.*, "Manifold regularized multitask feature learning for multimodality disease classification," *Human brain mapping*, vol. 36, no. 2, pp. 489–507, 2015.
- [32] B. T. Wyman *et al.*, "Standardization of analysis sets for reporting results from ADNI MRI data," *Alzheimer's & Dementia*, vol. 9, no. 3, pp. 332–337, 2013.
- [33] R. Wolz *et al.*, "Multi-method analysis of MRI images in early diagnostics of Alzheimer's disease," *PLoS one*, vol. 6, no. 10, p. e25446, 2011.
- [34] C. R. Jack *et al.*, "The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods," *Journal of Magnetic Resonance Imaging*, vol. 27, no. 4, pp. 685–691, 2008.
- [35] K. K. Leung *et al.*, "Brain MAPS: an automated, accurate and robust brain extraction technique using a template library," *NeuroImage*, vol. 55, no. 3, pp. 1091–1108, 2011.
- [36] D. Rueckert *et al.*, "Nonrigid registration using free-form deformations: application to breast MR images," *IEEE Transactions on Medical Imaging*, vol. 18, no. 8, pp. 712–721, 1999.
- [37] L. G. Nyu and J. K. Udupa, "On standardizing the MR image intensity scale," *Magnetic Resonance in Medicine*, vol. 42, no. 6, pp. 1072–1081, 1999.
- [38] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society*, pp. 267–288, 1996.
- [39] X. Zhu *et al.*, "A novel matrix-similarity based loss function for joint regression and classification in AD diagnosis," *NeuroImage*, vol. 100, pp. 91–105, 2014.
- [40] X. Zhu, H.-I. Suk *et al.*, "Subspace regularized sparse multi-task learning for multi-class neurodegenerative disease identification," *IEEE transaction on biomedical engineering*, 2015, in press.
- [41] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society*, vol. 67, no. 2, pp. 301–320, 2005.
- [42] N. Kriegeskorte *et al.*, "Circular analysis in systems neuroscience: the dangers of double dipping," *Nature neuroscience*, vol. 12, no. 5, pp. 535–540, 2009.
- [43] G. Wu *et al.*, "A generative probability model of joint label fusion for multi-atlas based brain segmentation," *Medical Image Analysis*, vol. 18, no. 6, pp. 881–890, 2014.
- [44] T. Tong *et al.*, "Segmentation of MR images via discriminative dictionary learning and sparse coding: Application to hippocampus labeling," *NeuroImage*, vol. 76, pp. 11–23, 2013.
- [45] J. Mairal *et al.*, "Online dictionary learning for sparse coding," in *International Conference on Machine Learning*, 2009, pp. 689–696.
- [46] J. D. Storey, "The positive false discovery rate: a Bayesian interpretation and the q-value," *Annals of statistics*, pp. 2013–2035, 2003.
- [47] C. Ledig *et al.*, "Alzheimer's disease state classification using structural volumetry, cortical thickness and intensity features," *MICCAI challenge on computer-aided diagnosis of dementia*, pp. 55–64, 2014.
- [48] R. Desikan *et al.*, "MRI measures of temporoparietal regions show differential rates of atrophy during prodromal AD," *Neurology*, vol. 71, no. 11, pp. 819–825, 2008.
- [49] R. I. Scahill *et al.*, "A longitudinal study of brain volume changes in normal aging using serial registered magnetic resonance imaging," *Archives of neurology*, vol. 60, no. 7, pp. 989–994, 2003.
- [50] N. C. Fox and J. M. Schott, "Imaging cerebral atrophy: normal ageing to Alzheimer's disease," *The Lancet*, vol. 363, no. 9406, pp. 392–394, 2004.
- [51] L. Ferrarini *et al.*, "Morphological hippocampal markers for automated detection of Alzheimer's disease and mild cognitive impairment converters in magnetic resonance images," *Journal of Alzheimer's Disease*, vol. 17, no. 3, pp. 643–659, 2009.
- [52] N. Singh *et al.*, "Genetic, structural and functional imaging biomarkers for early detection of conversion from MCI to AD," in *International Conference on MICCAI*, 2012, pp. 132–140.
- [53] A. Chincarini *et al.*, "Local MRI analysis approach in the diagnosis of early and prodromal Alzheimer's disease," *NeuroImage*, vol. 58, no. 2, pp. 469–480, 2011.
- [54] Y. Li *et al.*, "Discriminant analysis of longitudinal cortical thickness changes in Alzheimer's disease using dynamic and network features," *Neurobiology of Aging*, vol. 33, no. 2, pp. 427.e15–427.e30, 2012.
- [55] C.-Y. Wee *et al.*, "Prediction of Alzheimer's disease and mild cognitive impairment using cortical morphological patterns," *Human Brain Mapping*, vol. 34, no. 12, pp. 3411–3425, 2013.
- [56] E. Adeli-Mosabbeq *et al.*, "Robust feature-sample linear discriminant analysis for brain disorders diagnosis," in *Advances in Neural Information Processing Systems*, 2015, pp. 658–666.
- [57] O. Querbes *et al.*, "Early diagnosis of Alzheimer's disease using cortical thickness: impact of cognitive reserve," *Brain*, vol. 132, no. 8, pp. 2036–2047, 2009.
- [58] C. Davatzikos *et al.*, "Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification," *Neurobiology of Aging*, vol. 32, no. 12, pp. 2322.e19–2322.e27, 2011.
- [59] J. Koikkalainen *et al.*, "Multi-template tensor-based morphometry: application to analysis of Alzheimer's disease," *NeuroImage*, vol. 56, no. 3, pp. 1134–1144, 2011.
- [60] D. Zhang *et al.*, "Predicting future clinical changes of MCI patients using longitudinal and multimodal biomarkers," *PLoS one*, vol. 7, no. 3, p. e33182, 2012.