



# Classifying MCI Subtypes in Community-Dwelling Elderly Using Cross-Sectional and Longitudinal MRI-Based Biomarkers

Hao Guan<sup>1</sup>, Tao Liu<sup>1,2,3\*</sup>, Jiyang Jiang<sup>4,5</sup>, Dacheng Tao<sup>6,7</sup>, Jicong Zhang<sup>1,2</sup>, Haijun Niu<sup>1,3</sup>, Wanlin Zhu<sup>4,8</sup>, Yilong Wang<sup>8</sup>, Jian Cheng<sup>9\*</sup>, Nicole A. Kochan<sup>4,5</sup>, Henry Brodaty<sup>4,10</sup>, Perminder Sachdev<sup>4,5</sup> and Wei Wen<sup>4,5</sup>

<sup>1</sup> School of Biological Science and Medical Engineering, Beihang University, Beijing, China, <sup>2</sup> Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, Beijing, China, <sup>3</sup> Beijing Advanced Innovation Center for Biomedical Engineering, Beijing, China, <sup>4</sup> Centre for Healthy Brain Ageing, School of Psychiatry, University of New South Wales, Sydney, NSW, Australia, <sup>5</sup> Neuropsychiatric Institute, Prince of Wales Hospital, Sydney, NSW, Australia, <sup>6</sup> UBTech Sydney Artificial Intelligence Institute, Faculty of Engineering and Information Technologies, University of Sydney, Darlington, NSW, Australia, <sup>7</sup> The School of Information Technologies, Faculty of Engineering and Information Technologies, University of Sydney, Darlington, NSW, Australia, <sup>8</sup> Beijing Tiantan Hospital, Capital Medical University, Beijing, China, <sup>9</sup> NIBIB, NICHD, National Institutes of Health, Bethesda, MD, United States, <sup>10</sup> Dementia Collaborative Research Centre, University of New South Wales, Sydney, NSW, Australia

## OPEN ACCESS

### Edited by:

Javier Ramirez,  
University of Granada, Spain

### Reviewed by:

Iman Beheshti,  
National Center of Neurology and  
Psychiatry, Japan  
Heung-Il Suk,  
University of North Carolina at Chapel  
Hill, United States

### \*Correspondence:

Tao Liu  
tao.liu@buaa.edu.cn  
Jian Cheng  
jiancheng@ieee.org

**Received:** 20 July 2017

**Accepted:** 12 September 2017

**Published:** 26 September 2017

### Citation:

Guan H, Liu T, Jiang J, Tao D, Zhang J, Niu H, Zhu W, Wang Y, Cheng J, Kochan NA, Brodaty H, Sachdev P and Wen W (2017) Classifying MCI Subtypes in Community-Dwelling Elderly Using Cross-Sectional and Longitudinal MRI-Based Biomarkers. *Front. Aging Neurosci.* 9:309. doi: 10.3389/fnagi.2017.00309

Amnesic MCI (aMCI) and non-amnesic MCI (naMCI) are considered to differ in etiology and outcome. Accurately classifying MCI into meaningful subtypes would enable early intervention with targeted treatment. In this study, we employed structural magnetic resonance imaging (MRI) for MCI subtype classification. This was carried out in a sample of 184 community-dwelling individuals (aged 73–85 years). Cortical surface based measurements were computed from longitudinal and cross-sectional scans. By introducing a feature selection algorithm, we identified a set of discriminative features, and further investigated the temporal patterns of these features. A voting classifier was trained and evaluated via 10 iterations of cross-validation. The best classification accuracies achieved were: 77% (naMCI vs. aMCI), 81% (aMCI vs. cognitively normal (CN)) and 70% (naMCI vs. CN). The best results for differentiating aMCI from naMCI were achieved with baseline features. Hippocampus, amygdala and frontal pole were found to be most discriminative for classifying MCI subtypes. Additionally, we observed the dynamics of classification of several MRI biomarkers. Learning the dynamics of atrophy may aid in the development of better biomarkers, as it may track the progression of cognitive impairment.

**Keywords:** mild cognitive impairment, longitudinal data, early diagnosis, MRI, biomarker, feature selection, machine learning

## INTRODUCTION

Mild cognitive impairment (MCI) is thought to be a transitional stage between cognitively normal and dementia (Petersen, 2004). Previous studies have shown that neuroimaging biomarkers are potential predictors of cognitive impairment (Shi et al., 2010; Cuingnet et al., 2011; Davatzikos et al., 2011; Falahati et al., 2014; Trzepacz et al., 2014; Bron et al., 2015; Jung et al., 2016; Lebedeva et al., 2017). Many researchers have developed and implemented machine learning systems which

use neuroimaging biomarkers for more accurate identification of individuals with MCI or dementia (Cui et al., 2012a; Shao et al., 2012; Lebedev et al., 2014; Min et al., 2014; Moradi et al., 2015; Yun et al., 2015; Cai et al., 2017; Guo et al., 2017). Early diagnosis is an essential step in the prevention and early treatment of MCI and dementia.

MCI is clinically heterogeneous with different risks of progression to dementia. Clinical subtypes of MCI have been proposed to broaden the concept, and included prodromal forms of a variety of dementias (Petersen, 2004). MCI is termed “amnesic MCI” (aMCI) when memory loss is the predominant symptom. Almost 10% to 15% aMCI individuals tend to progress to clinically probable Alzheimer’s disease (AD) annually (Grundman et al., 2004). Additionally, MCI is termed “non-amnesic MCI” (naMCI) when impairments are in domains other than memory. Individuals with naMCI were more likely to convert to dementia other than AD, such as vascular dementia or dementia with Lewy bodies (Tabert et al., 2006). The progression of different MCI subtypes to a particular type of dementia has yet to be clearly delineated. On the other hand, MCI does not necessarily lead to dementia, since some studies suggested that MCI subjects have higher rates of reversion to normal cognition than progression to dementia (Brodaty et al., 2013; Pandya et al., 2016). A population-based study found that the reversion rate is lower in aMCI compared with naMCI (Roberts et al., 2014). Reliably identifying MCI of different subtypes would enable more efficient clinical trials and facilitate better targeted treatments.

Longitudinal measurements of Magnetic Resonance Imaging (MRI) in MCI and dementia may provide crucial predictors for tracking the disease progression of dementia (Misra et al., 2009; Risacher et al., 2010; Liu et al., 2013; Mayo et al., 2017). However, only a few studies used longitudinal data for automated classification of MCI and dementia (McEvoy et al., 2011; Li et al., 2012; Zhang et al., 2012a; Ardekani et al., 2017; Huang et al., 2017). Zhang et al. proposed an AD prediction method using longitudinal data which achieved greater classification results than using baseline visit data (Zhang et al., 2012a). Huang et al. presented a longitudinal measurement of MCI brain images and a hierarchical classification method for AD prediction. Their method using longitudinal data consistently outperformed the method using baseline data only (Huang et al., 2017). Despite these efforts, employing machine learning technique with longitudinal MRI features for MCI subtypes classification is rarely studied. And an additional aspect of research when using longitudinal MRI measurements is to identify the biomarkers that remain significant during the time course.

In this study, we used machine learning technique to classify MCI subtypes by employing cross-sectional and longitudinal MRI features. We reported nine independent classification experiments, whereby we compared two groups in each experiment: aMCI vs. cognitively normal (CN), naMCI vs. CN, naMCI vs. aMCI, using features measured at baseline, two-year follow-up, and longitudinally. The longitudinal features were employed by calculating the means and changes of the cross-sectional measurements. Clinical classifications at two-year follow-up were used as the comparison. The features used for classification were cortical surface based, including sulcal width,

cortical thickness, cortical gray matter (GM) volume, subcortical volumes and white matter hyper-intensity (WMH) volume. We compared the classification performance using cross-sectional features and longitudinal features. In addition, we performed feature selection and analyzed the temporal patterns of the selected biomarkers.

## MATERIALS AND METHODS

### Participants

Participants were members of the Sydney Memory and Aging Study (MAS), a longitudinal study of community-dwelling individuals aged 70–90 years recruited via the electoral roll from two regions of Sydney, Australia (Sachdev et al., 2010). Individuals were excluded at baseline if they had a previous diagnosis of dementia, mental retardation, psychotic disorder including schizophrenia or bipolar disorder, multiple sclerosis, motor neuron disease, developmental disability, or progressive malignancy. The study was approved by the Ethics Committees of the University of New South Wales and the South Eastern Sydney and Illawarra Area Health Service. Written informed consent was obtained from each participant.

### Diagnosis

Participants were diagnosed with MCI using the international consensus criteria (Winblad et al., 2004). Specifically, the presence of cognitive impairment as determined by performance on a neuropsychological measure of at least 1.5 standard deviations below published normative values for age and/or education on a test battery covering five cognitive domains (memory, attention/information processing, language, spatial and executive abilities), a subjective complaint of decline in memory or other cognitive function either from the participant or informant, and normal or minimally impaired instrumental activities of daily living attributable to cognitive impairment (total average score <3.0 on the Bayer Activity of Daily Living Scale, Hindmarch et al., 1998).

MCI were classified into two subtypes (aMCI or naMCI) according to cognitive impairment profiles (Petersen, 2004). Participants with no impairments on neuropsychological tests were deemed to have normal cognition. In this study, we included individuals who had MRI scans from both baseline and 2-year follow-up (wave-2), and a wave-2 diagnosis of either cognitively normal or MCI. Demographic characteristics were detailed in **Table 1**. A total of 184 participants met these criteria, including 115 cognitively normal (CN), 42 aMCI, and 27 naMCI. The MRI measurements used in the present study have been previously published (Liu et al., 2013).

### Image Acquisition

MRI scans were obtained with a 3-T system (Philips Medical Systems, Best, The Netherlands) using the same sequence for both baseline and follow-up scans:  $TR = 6.39$  ms,  $TE = 2.9$  ms, flip angle =  $8^\circ$ , matrix size =  $256 \times 256$ , FOV =  $256 \times 256 \times 190$  mm, and slice thickness = 1 mm with no gap, yielding  $1 \times 1 \times 1$  mm<sup>3</sup> isotropic voxels.

**TABLE 1** | Demographic characteristics of the sample.

Time point	Diagnostic group	No. of subjects (male)	Age mean (SD)	Years of Edu mean (SD)	MMSE score mean (SD)
Baseline	Total	184 (91)	77.48 (4.40)	11.79 (3.60)	28.16 (1.32)
	CN	117 (56)	77.12 (4.43)	11.93 (3.53)	28.39 (1.23)
	aMCI	40 (28)	78.36 (4.11)	11.81 (3.96)	27.58 (1.32)
	naMCI	27 (7)	77.76 (4.65)	11.14 (3.42)	28.00 (1.44)
Wave-2	Total	184 (91)	79.38 (4.40)	11.79 (3.60)	28.40 (1.41)
	CN	115 (53)	78.78 (4.15)	12.06 (3.42)	28.83 (1.16)
	aMCI	42 (30)	81.26 (4.98)	11.87 (4.16)	27.64 (1.59)
	naMCI	27 (8)	79.03 (3.72)	10.49 (3.27)	27.78 (1.40)

CN, cognitively normal; aMCI, amnesic mild cognitive impairment (MCI); naMCI, non-amnesic MCI; Edu, education, MMSE, Mini-mental state examination.

## Image Processing Sulcal Measures

Cortical sulci were extracted from the images via the following steps. First, non-brain tissues were removed to produce images containing only GM, white matter (WM) and cerebrospinal fluid (CSF). This was done by warping a brain mask defined in the standard space back to the T1-weighted structural MRI scan. The brain mask was obtained with an automated skull stripping procedure based on the SPM5 skull-cleanup tool (Ashburner, 2009). Individual sulci were identified and extracted using the BrainVisa (BV, version 3.2) sulcal identification pipeline (Rivière et al., 2009). A sulcal labeling tool incorporating 500 artificial neural network-based pattern classifiers (Rivière et al., 2002; Sun et al., 2007) was used to label sulci. Sulci that were mislabeled by BV were manually corrected. For each hemisphere, we determined the average sulcal width for five sulci: superior frontal, intra-parietal, superior temporal, central, and the sylvian fissure. Sulcal width was defined as the average 3D distance between opposing gyral banks along the normal projections to the medial sulcal mesh (Kochunov et al., 2012). The five sulci investigated in the present study were chosen because they were present in all individuals, large and relatively easy to identify after facilitating error detection and correction, and located on different cerebral lobes. For each hemisphere, we calculated the global sulcal index (g-SI) as the ratio between the total sulcal area and outer cortical area (Penttilä et al., 2009). We calculated the g-SI of each brain with no manual intervention using BV.

## Cortical Thickness, GM Volume

We computed average regional GM volume, average regional cortical thickness using the longitudinal stream in FreeSurfer 5.1 (<http://surfer.nmr.mgh.harvard.edu/>) (Reuter et al., 2012). This stream specifically creates an unbiased specific within-subject template space and image using robust, inverse consistent registration (Reuter and Fischl, 2011; Reuter et al., 2012). Briefly, this pipeline included the following processing steps, skull stripping, Talairach transforms, atlas registration, spherical surface maps, and parcellation of cerebral cortex (Desikan et al., 2006; Reuter et al., 2012). We applied Desikan parcellation (Desikan et al., 2006) which resulted 34 cortical regions of interest (ROIs) in each hemisphere. We visually inspected registration

and segmentation. Scans were excluded if they failed visual quality control, resulting in an unequal number of scans available for different brain structures. We calculated both the cortical thickness and the regional volumes for every cortical regions of the Desikan parcellation.

## Subcortical Volume

Subcortical brain structures were extracted using FSL's FIRST (FMRIB Image Registration and Segmentation Tool, Version 1.2), a model-based segmentation/registration tool (Patenaude et al., 2011). We included the following left and right subcortical structures: thalamus, caudate, putamen, pallidum, hippocampus, amygdala, and nucleus accumbens. Briefly, the FIRST algorithm modeled each participant's subcortical structure as a surface mesh, using a Bayesian model incorporating a training set of all images. We conducted visual quality control of FSL results using ENIGMA protocols (<http://enigma.ini.usc.edu/>). Three slices of each of coronal, sagittal and axial planes were extracted from each linearly transformed brain. For comparison, an outline of the templates was mapped onto the slices. We confirmed that the size of the participant brain corresponded with that of the template, verified that the lobes were appropriately situated, and confirmed that the orientation of the participant matched the template.

## WMHs

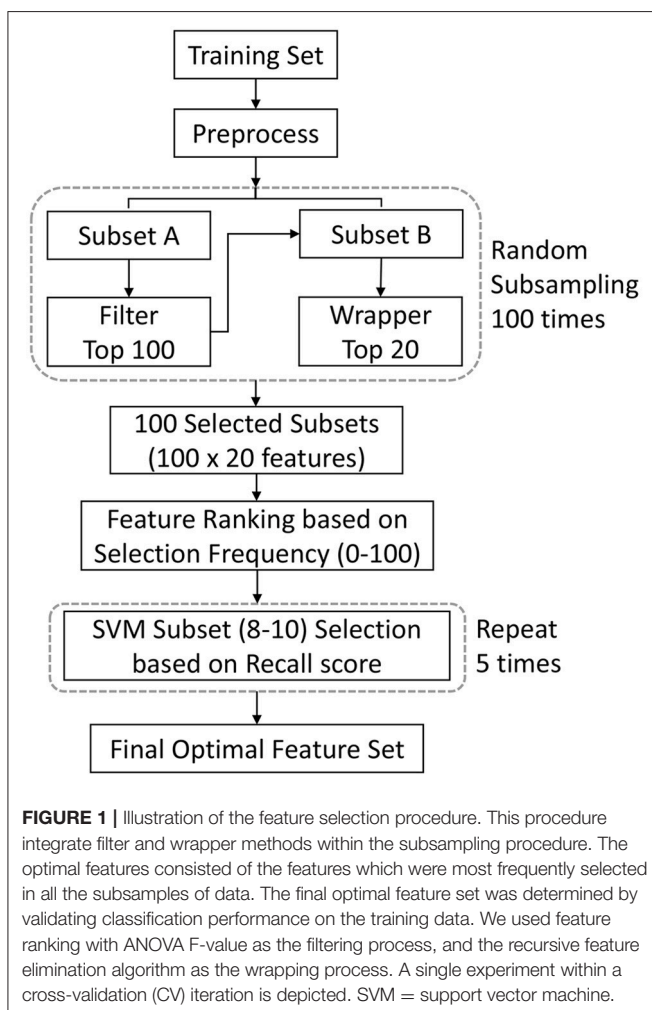
WMHs were delineated from coronal plane 3D T1-weighted and Fluid Attenuated Inversion Recovery (FLAIR) structural image scans using a pipeline described in detail previously (Wen et al., 2009). For each hemisphere, we calculated WMH volumes of eight brain regions: temporal, frontal, occipital, parietal, ventricle body, anterior horn, posterior horn, and cerebellum.

We obtained neuroimaging measurements of all participants at baseline and wave-2. The changes and the means values of those measurements were considered as the longitudinal features. There were altogether 178 MRI measurements for baseline and wave-2 feature sets, which included 12 sulcal measurements, 68 thickness measurements, 68 volume measurements, 14 subcortical measurements, and 16 WMH measurements. With the means and the changes, the longitudinal feature set included 356 MRI measurements.

## Feature Selection

The aims of feature selection were to maximize the performance of classification by identifying the most discriminative features, and help in understanding the neuropathological basis of neurocognitive impairments such as MCI and dementia. Supervised feature selection methods were often divided into three categories, namely “filter,” “wrapper,” and “embedded,” respectively (Mwangi et al., 2014). A particular problem of those methods was that when they were applied in the neuroimaging fields, where the number of features largely exceeded the number of examples, the cross-validation based error estimates usually led to results with extremely large variances (Dougherty et al., 2010; Tohka et al., 2016). We proposed a feature selection method in this study to reduce the variances by integrating the filter and the wrapper procedures within the subsampling iterations. The optimal feature subset consisted of the features which were most frequently selected in all the subsamples of data. The discriminative abilities of the features were assessed in terms of the selection frequencies.

**Figure 1** shows the flowchart of the feature selection procedure used in our study. We first randomly subsampled the training set 100 times. During each subsampling iteration,



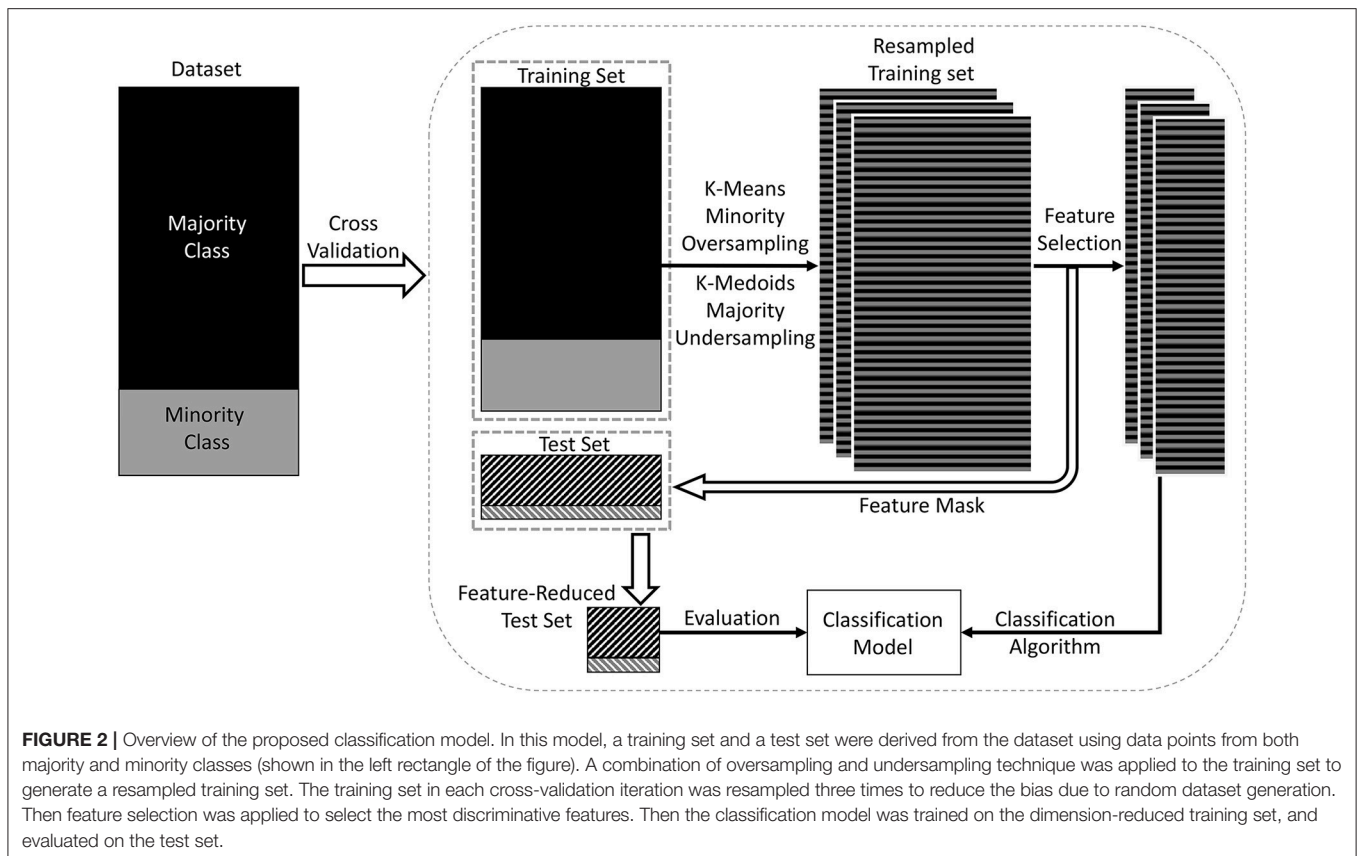
data were divided into two subsets of equal size, subset A and subset B. Subset A was processed by a filter to select features. The selected features were then applied to subset B. The subset B was processed by a wrapper to further reduce the number of features. After the subsampling processes, features were subsequently ranked in order of selection frequencies. The final optimal feature set was then determined by validating classification performance on the training data, using features chosen on the basis of frequency rank thresholds.

In the filter stage, ANOVA (analysis of variance) *F*-value were used to rank features on the basis of correlations with their diagnostic label. The top 100 features were selected at this stage. Then in the wrapping stage, the recursive feature elimination algorithm (Guyon et al., 2002) was used to further remove less informative features. Among the top 100 features, 20 were retained in this stage. The selection frequencies could be 100 at maximum or 0 at minimum. To mitigate the curse-of-dimensionality problem, the final feature set was limited with less than 10 features, and a variation section was established for the feature set to achieve the best validation performance. Given a frequency rank threshold  $N_f$  ( $N_f \in [10, 9, 8]$ ), we randomly split the training data into 2 subgroups: one for training a SVM (Vapnik, 1995) classifier with top  $N_f$  features, and the other for validation. The kernel for the SVM is the radial basic function (rbf). This step was repeated 5 times, and the recall scores were computed (the recall score is the ratio  $T_p/(T_p + F_n)$ , where  $T_p$  is the number of true positives and  $F_n$  is the number of false negatives). We chose the recall score as the criteria to minimize the impact of sample proportion imbalance. The top  $N_f$  features with the highest average recall score became the optimal feature set. We also evaluated the selected features using 2-tailed *t*-test.

## Classification and Validation

The imbalance of the sample could lead to a suboptimal classification performance. This study investigated a population-based sample, consisting of more cognitively normal individuals than MCI. There was also a large difference between the sample sizes of different MCI subtypes. We addressed this problem by using the data-resampling technique (Chawla et al., 2002; Dubey et al., 2014). An overview of the procedure is shown in **Figure 2**. We used a combination of oversampling and undersampling (Batista et al., 2004). K-means clustering (Macqueen, 1967) algorithm was used for oversampling, where new synthetic data were generated by clustering the minority class data. Briefly,  $N_s$  samples were clustered into  $N_s/3$  clusters, and  $N_s/3$  centroids were generated. Then these centroids and the original samples were combined for the next iteration of oversampling. The oversampling procedure was repeated until the size of minority class was  $2/3$  the size of the majority class. K-Medoids clustering (Hastie et al., 2001) algorithm was used for undersampling, where actual data points from the majority class were chosen as the cluster centers. The final training set was a combination of the oversampled minority class data and the undersampled majority class data. While resampling the training set, the test set remained the same. The training set was resampled 3 times to reduce the bias due to random data generation. Then the





feature selection method was applied on those resampled training sets, thus producing 3 learning models. These models were combined using majority voting, where the final label of an instance was decided based on the majority votes received from all the models.

We chose Voting Classifier for classification (Maclin and Opitz, 1999). A Voting Classifier combines conceptually different machine learning classifiers and uses a majority vote or the average predicted probabilities (soft vote) to predict the class labels. The advantage of Voting Classifier is to balance out the individual weaknesses of a set of equally well performing models. We chose SVM (rbf kernel), Logistic Regression (LR) (Cox, 1958), and Random Forest (RF) (Breiman, 2001) as the estimators of the Voting Classifier. All the estimators were with default settings of parameters. Specific weights (1:4:1) were assigned to SVM, LR and RF via the weights parameter. The weights were selected experimentally to aim at a better sensitivity score. We started with the equal weights (1:1:1), and changed the weights to obtain the best results. The predicted class probabilities of each classifier were collected, multiplied by the weights of classifiers, and averaged. The final class label was then derived from the class label with the highest average probability. As different features had different scales, we standardized all the training data within a 0–1 range, and the same procedure was then applied to the test data.

We evaluated our method using stratified Shuffle Split cross-validation procedure, also known as Monte Carlo cross-validation (Berrar et al., 2007), which returned stratified randomized folds by preserving the percentage of samples for each class. The cross-validation procedure was repeated 10 times with a fixed 9:1 train-test ratio. The final classification results represented the average of these 10 independent experiments. We applied four metrics to assess the performance of the model: the accuracy, the specificity, the sensitivity, and the area under the receiver operating characteristic curve (AUC). AUC is a better measure than accuracy in imbalanced data sets and real-world applications (Huang and Ling, 2005; Bekkar et al., 2013).

It was important to note that we obtained a unique set of selected features in each training set. The training set in each cross-validation iteration was resampled 3 times, thus producing 3 resampled training sets. In each training set, the maximum possible selection frequency of one feature was 100. Considering the feature selection and data-resampling steps within the 10-iteration cross-validation procedure, the final maximum possible selection frequency of each feature was  $3 \times 100 \times 10 = 3,000$ .

All the data processing and analyzing were performed using Python libraries Numpy 1.10.4 (Walt et al., 2011) and Scipy 0.17.0 (Jones et al., 2001) on Python 2.7.11 (Anaconda 4.0.0–64 bit, <http://www.continuum.io/>). All the machine learning methods were performed using the library Scikit-Learn 0.17.1 (Pedregosa et al., 2011).

## RESULTS

### MCI Subtypes Classification

As shown in **Table 2**, in the classification of aMCI and CN, compared with using baseline features, using longitudinal features improved the performance to accuracy of 73%, sensitivity of 53%, specificity of 80%, and AUC of 0.75; the results of using longitudinal features were not superior to that using wave-2 features. Identifying naMCI from CN was relatively difficult considering the poor sensitivity value and AUC; the results of using longitudinal and cross-sectional features were comparable and without significant difference. In the classification of naMCI vs. aMCI, compared with using longitudinal features, using baseline features achieved better performance; the results of using wave-2 features were not significantly different from using longitudinal features.

### Discriminative Features

The discriminative ability of the features used in this study were assessed by examining the frequency with which they were selected. We listed the top 10 most frequently selected features in each MCI subtype classification experiment (see **Tables 3–5**). In the comparison of aMCI vs. CN, thickness of right frontal pole, left superior temporal, volume of right thalamus, and right hippocampus were more discriminative than the rest of features (see **Table 3**). In the classification of naMCI vs. aMCI, thickness of right rostral middle frontal, right pericalcarine, right frontal pole, and volume of right rostral anterior cingulate were more discriminative than the others (see **Table 5**). Regardless of cross-sectional (baseline and wave-2) or longitudinal, all the features mentioned above were listed in the top-10 feature list. In the naMCI vs. CN comparison, volume of left temporal pole and right amygdala were also discriminative (see **Table 4**).

The top-10 selected features were analyzed to identify the temporal patterns. Several features measured at different time points showed dynamic discriminative powers. **Figures 3–5** shows the selection frequencies of the stable features measured at each time point. A feature may be identified as stable when this

feature was selected at all the baseline, wave-2, and longitudinally. The selection frequencies of the stable features for aMCI vs. CN classification are shown in **Figure 3**. We observed that thickness of right frontal pole was a stable biomarker, since its selection frequencies were close between different time points. The selection frequencies of several biomarkers changed visibly over time, including volume of right thalamus, right hippocampus, and thickness of left superior temporal. In the classification of naMCI vs. CN (see **Figure 4**), only a few features were stable. We observed that the volume of right amygdala provided more useful information at baseline. Volume of left temporal pole and right rostral cingulate carried more information at baseline. In the classification of naMCI vs. aMCI (see **Figure 5**), volume of right rostral middle frontal and thickness of right pericalcarine thickness were selected more often at baseline, while volume of right frontal pole were more discriminative at wave-2. And volume of right rostral anterior cingulate provided important information at all-time points.

Furthermore, some features were selected in the top-10 feature list at either baseline or wave-2, such as the right g-SI index, sulcal width of superior frontal (see **Table 3**); thickness of left lateral occipital, WMH volume of right cerebellum (see **Table 4**); thickness of right lateral occipital, and WMH volume of right frontal (see **Table 5**). On the other hand, some features were selected only in longitudinal cases, such as sulcal width of right superior temporal, thickness of left inferior temporal (see **Table 3**); volume of right entorhinal and right posterior cingulate, thickness of left posterior cingulate and temporal pole (see **Table 4**); thickness of left precentral, volume of right entorhinal (see **Table 5**). Most of these longitudinal features were the differences (changes value) between the measures of two time points.

## DISCUSSION

Our study examined classification of MCI subtypes in community-dwelling elderly using cross-sectional and

**TABLE 2** | Classification results of MCI subtypes: features measured at baseline, wave-2 and longitudinally are used and compared.

Task	No. of minority class	No. of majority class	Method	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
aMCI vs. CN	aMCI = 42	CN = 115	Baseline	0.64	0.42	0.71	0.68
			Wave-2	0.81*	0.68	0.85	0.74
			Longitudinal	0.73	0.53	0.80	0.75
naMCI vs. CN	naMCI = 27	CN = 115	Baseline	0.67	0.37	0.75	0.57
			Wave-2	0.65	0.30	0.74	0.58
			Longitudinal	0.70	0.23	0.82	0.60
naMCI vs. aMCI	naMCI = 27	aMCI = 42	Baseline	0.77*	0.70*	0.82	0.84*
			Wave-2	0.71	0.57	0.82	0.70
			Longitudinal	0.61	0.40	0.78	0.71

wave-2, 2-year follow-up; MCI, mild cognitive impairment; CN cognitively normal; aMCI, amnesic MCI; naMCI, non-amnesic MCI; AUC, area under the receiver operating characteristic curve.

\*Significantly different from the method using longitudinal features; results are from *t*-test ( $p < 0.05$ ).

**TABLE 3** | Selected features for the classification of aMCI vs. CN.

Task	Baseline feature	Frequency	p <sup>a</sup>	Wave-2 feature	Frequency	p <sup>a</sup>	Longitudinal feature	Frequency	p <sup>a</sup>
<b>aMCI vs. CN</b>	<b>Right frontal pole thickness</b>	<b>2,658</b>	<b>0.001</b>	<b>Right frontal pole thickness</b>	<b>2,786</b>	<b>&lt;0.001</b>	<b>Right frontal pole thickness</b>	<b>2,715</b>	<b>&lt;0.001</b>
	<b>Right thalamus volume</b>	<b>2,136</b>	<b>0.006</b>	<b>Right hippocampus volume</b>	<b>2,105</b>	<b>&lt;0.001</b>	<b>Right thalamus volume</b>	<b>2,470</b>	<b>0.001</b>
	<b>Left superior temporal thickness</b>	<b>1,620</b>	<b>0.001</b>	<b>Right thalamus volume</b>	<b>1,775</b>	<b>0.001</b>	<b>Right hippocampus volume</b>	<b>2,071</b>	<b>0.001</b>
	Right hippocampus volume	1,344	0.005	Left superior temporal thickness	1,144	0.005	Left superior temporal thickness	1,212	0.002
	Right g-SI <sup>c</sup>	1,265	0.003	Left sulcal width of superior frontal	1,062	0.002	Right sulcal width of superior temporal <sup>b*</sup>	1,036	0.027
	Right transverse temporal thickness <sup>c</sup>	755	0.013	Right sulcal width of superior frontal <sup>c</sup>	963	0.001	Right pericalcarine thickness	876	0.004
	Right pericalcarine thickness	736	0.005	Right amygdala volume <sup>c</sup>	963	0.073	Left precentral thickness <sup>b*</sup>	869	0.035
	Right rostral anterior cingulate volume <sup>c</sup>	693	0.128	Right pericalcarine thickness	958	0.006	Left inferior temporal thickness*	827	0.019
	Right paracentral thickness <sup>c</sup>	637	0.022	Right accumbens volume <sup>c</sup>	660	0.011	Right paracentral thickness <sup>b*</sup>	819	0.012
	Left posterior cingulate volume <sup>c</sup>	545	0.134	Left medial orbitofrontal thickness <sup>c</sup>	633	0.045	Right sulcal width of superior frontal	722	0.002

A feature measured at baseline, wave-2 or longitudinally is defined as baseline feature, wave-2 feature or longitudinal feature, respectively. The first 10 most frequently selected features and their selection frequencies are listed. The maximum possible selection frequency of each feature is 3000. The features with selection frequencies above 1500 are in bold. wave-2, 2-year follow-up; MCI, mild cognitive impairment; CN, cognitively normal; aMCI, amnesic MCI.

<sup>a</sup>Results for comparisons of positive subjects and negative subjects using *t*-tests.

<sup>b</sup>Changes measurements, the rest longitudinal features are means measurements.

<sup>c</sup>Features that were selected at a single time point (either at baseline or wave-2).

\*Features that were selected only in longitudinal case.

longitudinal MRI measurements. Our classification framework implemented a data-resampling step to reduce the effect of the class-imbalance, and a feature selection step in which maximally most discriminative feature subsets were identified. The results suggested that individuals with aMCI could be differentiated from CN and naMCI with MRI-based biomarkers, but identifying naMCI from CN was still a challenge. Identifying aMCI from CN using longitudinal features achieved better performance than that using baseline features, but the results were not superior to that using wave-2 features. The best performance of differentiating aMCI from naMCI was achieved with baseline features. In addition, we analyzed and identified the dynamics of the biomarkers.

The subtlety of brain changes in MCI challenges the image-based classification. Previous studies reported using machine learning to differentiate MCI from cognitively normal (Wee et al., 2011, 2012; Zhang et al., 2011, 2017; Cui et al., 2012b; Liu et al., 2015, 2017). Cui et al. used combined measurements of T1-weighted and diffusion tensor imaging (DTI) to distinguish aMCI from CN, achieved a classification accuracy of 71%, sensitivity 52%, specificity 78%, and AUC 0.70 (Cui et al., 2012b). Our performance (accuracy 81%, sensitivity 68%, specificity 85%, and AUC 0.74) is better than their study. The approach of Wee et al. was a kernel combination method that utilized DTI and resting-state functional magnetic resonance imaging (Wee et al., 2012). Although their classification accuracy of 96.3% is higher than ours, the inclusion of multi-modality imaging could

restrict their use in clinical settings, and the small sample size of fewer than 30 participants may also make their results less robust. Considering the heterogeneity of MCI, we performed MCI subtypes classification, and the results demonstrated that aMCI and naMCI could be accurately separated with MRI biomarkers. And the results showed that the various groups demonstrated different patterns of atrophy on MRI. However, differentiating naMCI from CN was difficult considering the low sensitivities (see **Table 2**). The serious imbalance of classes could result in this poor performance, although we had performed data-resampling to mitigate the difference of the sample sizes. Compared with aMCI, naMCI individuals are more likely to revert to normal cognition (Roberts et al., 2014; Aerts et al., 2017). The MCI individuals who reverted might have different underlying mechanisms (Zhang et al., 2012b). In addition, higher estimates of MCI incidence in clinic-based studies (Petersen, 2004, 2010) than in population-based studies suggested that the rate of reversion to normal cognition may be lower in the clinic setting than in population-based studies (Koepsell and Monsell, 2011; Lopez et al., 2012) such as ours.

Longitudinal patterns of atrophy identified in MRI measurements can be used to elevate the prediction of cognitive decline (Rusinek et al., 2003; Risacher et al., 2010). McEvoy et al. investigated whether single-time-point and longitudinal volumetric MRI measures provided predictive prognostic information in patients with aMCI. Their results showed that the information regarding the rate of atrophy progression

**TABLE 4** | Selected features for the classification of naMCI vs. CN.

Task	Baseline feature	Frequency	p <sup>a</sup>	Wave-2 feature	Frequency	p <sup>a</sup>	Longitudinal feature	Frequency	p <sup>a</sup>
<i>naMCI vs. CN</i>	<b>Right WMH volume of cerebellum<sup>c</sup></b>	<b>2,240</b>	<b>0.014</b>	<b>Left lateral occipital thickness<sup>c</sup></b>	<b>2,087</b>	<b>0.002</b>	<b>Right entorhinal volume<sup>b*</sup></b>	<b>2,866</b>	<b>&lt;0.001</b>
	<b>Left temporal pole volume</b>	<b>2,227</b>	<b>0.072</b>	<b>Right rostral middle frontal thickness</b>	<b>1,670</b>	<b>0.024</b>	<b>Right amygdala volume</b>	<b>1,852</b>	<b>0.001</b>
	<b>Right amygdala volume</b>	<b>2,027</b>	<b>0.002</b>	<b>Left temporal pole volume</b>	<b>1,636</b>	<b>0.086</b>	<b>Right posterior cingulate volume<sup>b*</sup></b>	<b>1,608</b>	<b>0.008</b>
	<b>Right rostral middle frontal thickness</b>	<b>1,757</b>	<b>0.008</b>	<b>Right amygdala volume</b>	<b>1,527</b>	<b>0.003</b>	Left lateral occipital thickness	1,434	0.002
	<b>Right rostral anterior cingulate volume</b>	<b>1,718</b>	<b>0.011</b>	Right sulcal width of superior frontal <sup>c</sup>	1,259	0.017	Left temporal pole volume	1,256	0.074
	Left middle temporal thickness <sup>c</sup>	1,316	0.002	Left pericalcarine volume <sup>c</sup>	1,218	0.012	Left posterior cingulate thickness <sup>b*</sup>	891	0.022
	Right inferior parietal thickness <sup>c</sup>	953	0.002	Right rostral anterior cingulate volume	993	0.026	Left amygdala volume <sup>b*</sup>	746	0.117
	Right thalamus volume <sup>c</sup>	833	0.005	Right putamen volume <sup>c</sup>	754	0.001	Left temporal pole thickness <sup>b*</sup>	630	0.054
	left transverse temporal volume <sup>c</sup>	778	0.182	Right supramarginal volume	602	0.263	Left middle temporal thickness	613	0.007
	Right supramarginal volume	638	0.108	Left sulcal width of superior temporal <sup>c</sup>	515	0.134	Right WMH volume of cerebellum	578	0.378

A feature measured at baseline, wave-2 or longitudinally is defined as baseline feature, wave-2 feature or longitudinal feature, respectively. The first 10 most frequently selected features and their selection frequencies are listed. The maximum possible selection frequency of each feature is 3000. The features with selection frequencies above 1,500 are in bold. wave-2, 2-year follow-up; CN, cognitively normal; naMCI, non-amnesic MCI.

<sup>a</sup>Results for comparisons of positive subjects and negative subjects using *t*-tests.

<sup>b</sup>Changes measurements, the rest longitudinal features are means measurements.

<sup>c</sup>Features that were selected at a single time point (either at baseline or wave-2).

\*Features that were selected only in longitudinal case.

over a 1-year period improved risk prediction compared with using single-time-point MRI measurement (McEvoy et al., 2011). Huang et al. used longitudinal changes over 4 years of T1-weighted MRI scans to predict AD conversion in MCI subjects. Their results showed that the model with longitudinal data consistently outperformed the model with baseline data, especially achieved 17% higher sensitivity than the model with baseline data (Huang et al., 2017). In our study, the results showed that the longitudinal features failed to provide additional information for identifying aMCI and naMCI compared with cross-sectional features. In the classification of aMCI vs. CN, the accuracy with longitudinal features was nearly 10% higher than the accuracy with baseline features, but was not superior to the accuracy with wave-2 features (Table 2). The performance of using longitudinal features was comparable to using cross-sectional features at baseline and wave-2 for distinguishing naMCI from CN. In addition, the highest performance of distinguishing naMCI from aMCI was achieved with baseline features (see Table 2). This might be because the progression of naMCI showed no coherent pattern of atrophy. The patterns of atrophy differ among aMCI and naMCI, and subjects with naMCI showed scattered patterns of gray matter loss without any particular focus (Whitwell et al., 2007). All the subjects of our study were community-dwelling. It was likely that the naMCI subjects had atrophy patterns closer to those of CN at baseline, but over the time the patterns progressed to more MCI-like at wave-2. Our results also indicated that features selected for identifying naMCI were unstable over time, which

might be because clinical classification of naMCI can be based on impairment individually or in combination across a range of non-amnesic cognitive domains (language, visuo-spatial, processing speed, or executive abilities).

Longitudinal research has observed the dynamics of biomarkers (Trojanowski et al., 2010; Sabuncu et al., 2011; Eskildsen et al., 2013; Zhou et al., 2013). Some features provided significant information at all-time points while some other features were shown to be useful at a specific time point. Eskildsen et al. demonstrated that prediction accuracies of conversion from MCI to AD can be improved by learning the atrophy patterns that were specific to the different stages of disease progression (Eskildsen et al., 2013). They found that medial temporal lobe structures were stable biomarkers across all stages. Hippocampus was not discriminative at 36 months prior to AD diagnosis, but was included in all prediction cases of later stages. In addition, biomarkers were mostly selected from the cingulate gyrus, which is well known to be affected in early AD (Eskildsen et al., 2013). Histological studies suggest that the integrity of entorhinal cortex is among the first affected, which is then only later followed by an atrophy of the hippocampus (Braak et al., 1993). In our study, we also found that volume of the right hippocampus was more discriminative at wave-2 (see Figure 3, Table 3), which would complemented the histological findings. Furthermore, the thalamic volume was discriminative and stable over time (see Figure 3, Table 3), which was consistent with a previous study that the structure and function of thalamus determined severity of cognitive impairment (Schoonheim



**TABLE 5** | Selected features for the classification of naMCI vs. aMCI.

Task	Baseline feature	Frequency	p <sup>a</sup>	Wave-2 feature	Frequency	p <sup>a</sup>	Longitudinal feature	Frequency	p <sup>a</sup>
naMCI vs. aMCI	<b>Right rostral middle frontal thickness</b>	<b>2,643</b>	<b>&lt;0.001</b>	<b>Right rostral anterior cingulate volume</b>	<b>2,754</b>	<b>&lt;0.001</b>	<b>Right rostral anterior cingulate volume</b>	<b>2,598</b>	<b>&lt;0.001</b>
	<b>Right rostral anterior cingulate volume</b>	<b>2,538</b>	<b>&lt;0.001</b>	<b>Right frontal pole thickness</b>	<b>2,634</b>	<b>&lt;0.001</b>	<b>Right rostral middle frontal thickness</b>	<b>2,502</b>	<b>&lt;0.001</b>
	<b>Right pericalcarine thickness</b>	<b>2,241</b>	<b>0.001</b>	<b>Right rostral middle frontal thickness</b>	<b>2,190</b>	<b>&lt;0.001</b>	<b>Right frontal pole thickness</b>	<b>2,478</b>	<b>&lt;0.001</b>
	<b>Right frontal pole thickness</b>	<b>1,815</b>	<b>&lt;0.001</b>	<b>Right pericalcarine thickness</b>	<b>1,551</b>	<b>0.005</b>	<b>Right pericalcarine thickness</b>	<b>1,830</b>	<b>0.002</b>
	<b>Right g-si<sup>c</sup></b>	<b>1,539</b>	<b>0.004</b>	Left transverse temporal volume <sup>c</sup>	1,131	0.028	Right lateral occipital thickness	1,062	0.001
	Right lateral occipital thickness <sup>c</sup>	1,023	<0.001	Right wmh volume of frontal <sup>c</sup>	1,071	0.122	Right entorhinal volume <sup>d*</sup>	1,029	0.010
	Right transverse temporal thickness <sup>c</sup>	750	0.014	Left rostral middle frontal volume <sup>c</sup>	813	0.037	Left transverse temporal thickness	867	0.009
	Left inferior temporal thickness <sup>c</sup>	687	<0.001	Right insula thickness <sup>c</sup>	678	0.037	Left inferior temporal thickness	666	0.001
	Right parsorbitalis thickness <sup>c</sup>	666	0.002	Right frontal pole volume <sup>c</sup>	573	0.054	Left precentral thickness*	639	0.001
	Right transverse temporal thickness <sup>c</sup>	480	0.011	Right sulcal width of superior temporal <sup>c</sup>	552	0.026	Right g-SI	591	0.011

A feature measured at baseline, wave-2 or longitudinally is defined as baseline feature, wave-2 feature or longitudinal feature, respectively.

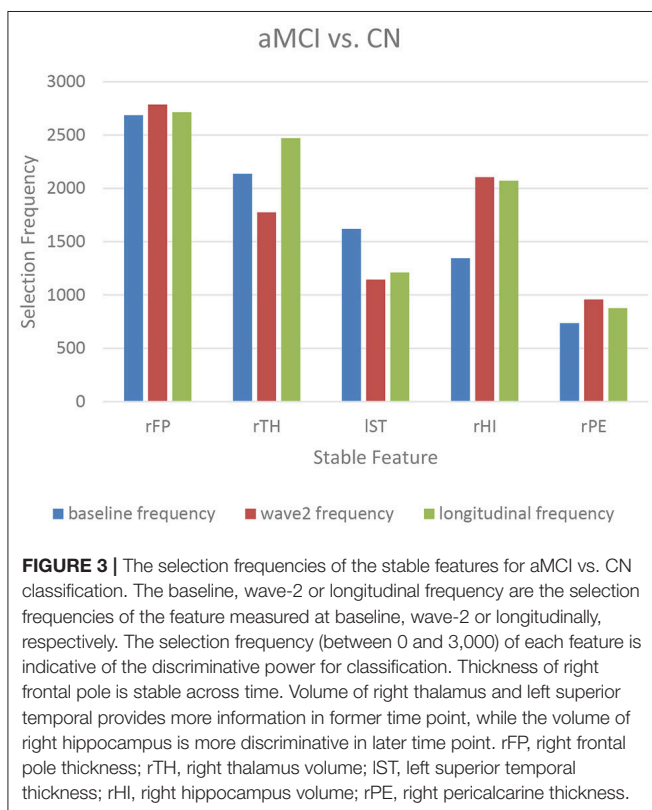
The first 10 most frequently selected features and their selection frequencies are listed. The maximum possible selection frequency of each feature is 3,000. The features with selection frequencies above 1500 are in bold. Key: wave-2, 2-year follow-up; aMCI, amnesic MCI; naMCI, non-amnesic MCI.

<sup>a</sup>Results for comparisons of positive subjects and negative subjects using t-tests.

<sup>b</sup>Change measurement, the rest longitudinal features are mean measurements.

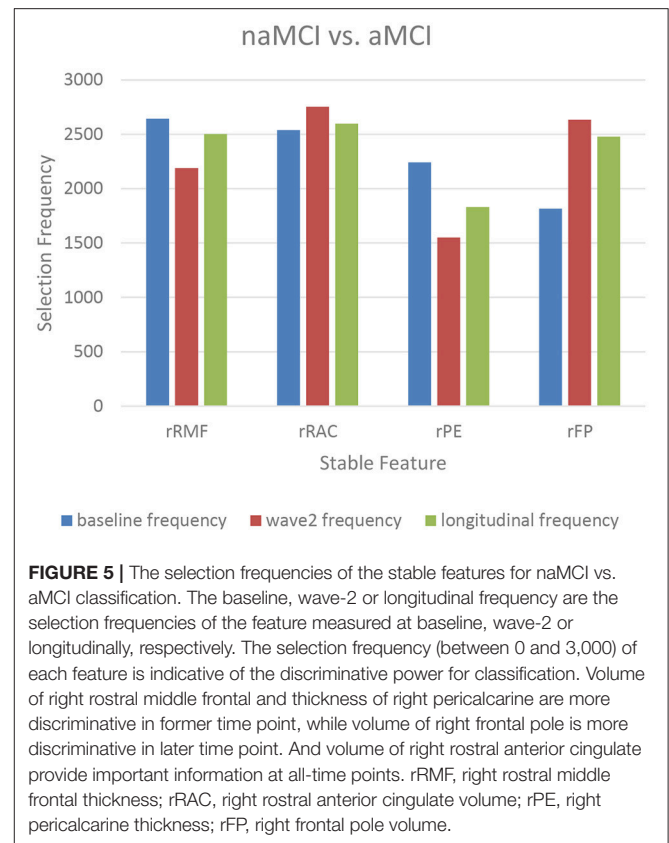
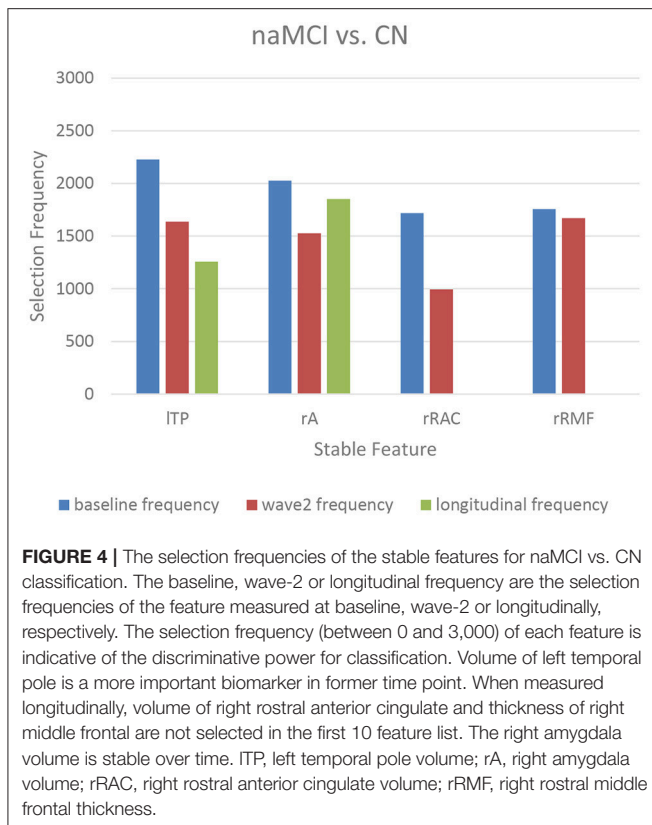
<sup>c</sup>Features that were selected at a single time point (either at baseline or wave-2).

\*Features that were selected only in longitudinal case.



et al., 2015). Volume of left posterior cingulate and right rostral anterior cingulate were more discriminative at baseline for identifying aMCI and naMCI from CN (see **Tables 3, 4**), while volume of right rostral anterior cingulate was a stable biomarker for naMCI vs. aMCI classification over time (see **Figure 5, Table 5**). Zhou et al. used the baseline MRI features to predict MMSE (The Mini-Mental State Examination, Folstein et al., 1975) and ADAS-Cog (Alzheimer's Disease Assessment Scale cognitive subscale, Rosen et al., 1984) scores in the next 4 years (Zhou et al., 2013). They observed that the average cortical thickness of left middle temporal, left and right entorhinal, and volume of left hippocampus were important biomarkers for predicting ADAS-Cog scores at all-time points. Cortical volume of left entorhinal provided significant information in later stages than in the first 6 months. Several biomarkers including volume of left and right amygdala provided useful information only at later time points (Zhou et al., 2013). In our study, cross-sectional (both baseline and wave-2) volume of right entorhinal was not an important biomarker for the classification of naMCI vs. CN, but the longitudinal volume change of right entorhinal (see **Table 4**) was discriminative. Volume of right amygdala was discriminative at all-time points for naMCI vs. CN classification (see **Figure 4, Table 4**). The dynamics of biomarker could potentially aid in developing stable imaging biomarkers and in tracking the progression of cognitive impairment.

The use of same dataset for feature selection and classification is termed "double-dipping," which will lead to distorted



descriptive statistics and artificially inflated accuracies (Kriegeskorte et al., 2009; Pereira et al., 2009; Eskildsen et al., 2013; Mwangi et al., 2014). Due to the limited samples in neuroimaging studies, carelessly designed training, testing and validation schemes, the risk of double-dipping is high. Eskildsen et al. used cortical regions potentially discriminative for predicting AD. They found that by inclusion of test subjects in the feature selection process, the prediction accuracies were artificially inflated (Eskildsen et al., 2013). In our experiments, training datasets and test datasets were adequately separated using cross-validation procedure. The training set in each cross-validation iteration were used for data-resampling, feature selection and classifier training, while the test set were only used for validating classification performance.

The main limitation of the present study was the limited sample size. Our method required longitudinal data, thus limiting the subjects with MRI scans at both time points. Secondly, this study investigated a population-based sample, consisting of more cognitively normal individuals than MCI. There was also a difference between the sample sizes of aMCI and naMCI. The findings need to be replicated in other data sets.

## CONCLUSION

In conclusion, the present study investigated MCI subtypes classification in a sample from community-dwelling elderly using both cross-sectional and longitudinal MRI features.

Our experiments suggested that longitudinal features were not superior to the cross-sectional features for MCI subtypes classifications. Dynamics of the biomarkers were analyzed and identified. Future studies with longer follow-up and more measurement occasions may lead to the better understanding of the trajectories for cognitive impairment.

## AUTHOR CONTRIBUTIONS

HG, TL, and JC: Study design, data analyses, interpretation of the results, manuscript writing. WW and DT: Study design, interpretation of the results. NK, PS, and HB: Data collection, interpretation of the results. JJ, JZ, HN, WZ, and YW: Data analyses. All authors participated in manuscript revision and final approval.

## FUNDING

This research received support from the Natural Science Foundation of China [grant numbers 81401476], the National Key Research and Development Program of China [grant numbers 2016YFF0201002], the National Health and Medical Research Council (NHMRC) Program Grants [grant numbers 350833, 56896, 109308], and the Australian Research Council Projects [grant numbers FL-170100117, DP-140102164, LP-150100671].

## REFERENCES

- Aerts, L., Heffernan, M., Kochan, N. A., Crawford, J. D., Draper, B., Trollor, J. N., et al. (2017). Effects of MCI subtype and reversion on progression to dementia in a community sample. *Neurology* 88, 2225–2232. doi: 10.1212/WNL.0000000000004015
- Ardekani, B. A., Bermudez, E., Mubeen, A. M., Bachman, A. H., and Alzheimer's Disease Neuroimaging Initiative (2017). Prediction of incipient Alzheimer's disease dementia in patients with mild cognitive impairment. *J. Alzheimers. Dis.* 55, 269–281. doi: 10.3233/JAD-160594
- Ashburner, J. (2009). Computational anatomy with the SPM software. *Magn. Reson. Imaging* 27, 1163–1174. doi: 10.1016/j.mri.2009.01.006
- Batista, G. E. A. P. A., Prati, R. C., and Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. *SIGKDD Explor. Newsl.* 6, 20–29. doi: 10.1145/1007730.1007735
- Bekkar, M., Djemaa, H. K., and Alitouche, T. A. (2013). Evaluation measures for models assessment over imbalanced data sets. *J. Inform. Eng. Appl.* 3, 27–38.
- Berrard, D., Granzow, M., and Dubitzky, W. (2007). "Introduction to genomic and proteomic data analysis," in *Fundamentals of Data Mining in Genomics and Proteomics* (Boston, MA: Springer).
- Braak, H., Braak, E., and Bohl, J. (1993). Staging of Alzheimer-related cortical destruction. *Eur. Neurol.* 33, 403–408. doi: 10.1159/000116984
- Breiman, L. (2001). Machine learning, volume 45, Number 1 - SpringerLink. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324
- Brodady, H., Heffernan, M., Kochan, N. A., Draper, B., Trollor, J. N., Reppermund, S., et al. (2013). Mild cognitive impairment in a community sample: the Sydney memory and ageing study. *Alzheimers Dement.* 9, 310–317. doi: 10.1016/j.jalz.2011.11.010
- Bron, E. E., Smits, M., van der Flier, W. M., Vrenken, H., Barkhof, F., Scheltens, P., et al. (2015). Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: the CADDementia challenge. *Neuroimage* 111, 562–579. doi: 10.1016/j.neuroimage.2015.01.048
- Cai, K. P., Xu, H., Guan, H., Zhu, W. L., Jiang, J. Y., Cui, Y., et al. (2017). Identification of early-stage Alzheimer's disease using sulcal morphology and other common neuroimaging indices. *PLoS ONE* 12:e0170875. doi: 10.1371/journal.pone.0170875
- Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *J. Artif. Intell. Res.* 16, 321–357. doi: 10.1613/jair.953
- Cox, D. R. (1958). The regression analysis of binary sequences. *J. R. Stat. Soc.* 20, 215–242.
- Cui, Y., Sachdev, P. S., Lipnicki, D. M., Jin, J. S., Luo, S., Zhu, W., et al. (2012a). Predicting the development of mild cognitive impairment: a new use of pattern recognition. *Neuroimage* 60, 894–901. doi: 10.1016/j.neuroimage.2012.01.084
- Cui, Y., Wen, W., Lipnicki, D. M., Beg, M. F., Jin, J. S., Luo, S., et al. (2012b). Automated detection of amnesic mild cognitive impairment in community-dwelling elderly adults: a combined spatial atrophy and white matter alteration approach. *Neuroimage* 59, 1209–1217. doi: 10.1016/j.neuroimage.2011.08.013
- Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehericy, S., Habert, M. O., et al. (2011). Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. *Neuroimage* 56, 766–781. doi: 10.1016/j.neuroimage.2010.06.013
- Davatzikos, C., Bhatt, P., Shaw, L. M., Batmanghelich, K. N., and Trojanowski, J. Q. (2011). Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. *Neurobiol. Aging* 32, 2322.e19–27. doi: 10.1016/j.neurobiolaging.2010.05.023
- Desikan, R. S., Segonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31, 968–980. doi: 10.1016/j.neuroimage.2006.01.021
- Dougherty, E. R., Sima, C., Hua, J. P., Hanczar, B., and Braga-Neto, U. M. (2010). Performance of error estimators for classification. *Curr. Bioinform.* 5, 53–67. doi: 10.2174/157489310790596385
- Dubey, R., Zhou, J. Y., Wang, Y. L., Thompson, P. M., Ye, J. P., and Alzheimer's Disease Neuroimaging Initiative. (2014). Analysis of sampling techniques for imbalanced data: an n=648 ADNI study. *Neuroimage* 87, 220–241. doi: 10.1016/j.neuroimage.2013.10.005
- Eskildsen, S. F., Coupe, P., Garcia-Lorenzo, D., Fonov, V., Pruessner, J. C., Collins, D. L., et al. (2013). Prediction of Alzheimer's disease in subjects with mild cognitive impairment from the ADNI cohort using patterns of cortical thinning. *Neuroimage* 65, 511–521. doi: 10.1016/j.neuroimage.2012.09.058
- Falahati, F., Westman, E., and Simmons, A. (2014). Multivariate Data Analysis and Machine Learning in Alzheimer's Disease with a Focus on Structural Magnetic Resonance Imaging. *J. Alzheimers Dis.* 41, 685–708. doi: 10.3233/JAD-131928
- Folstein, M., Folstein, S., and McHugh, P. (1975). "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* 12, 189–198. doi: 10.1016/0022-3956(75)90026-6
- Grundman, M., Petersen, R. C., Ferris, S. H., Thomas, R. G., Aisen, P. S., Bennett, D. A., et al. (2004). Mild cognitive impairment can be distinguished from Alzheimer disease and normal aging for clinical trials. *Arch. Neurol.* 61, 59–66. doi: 10.1001/archneur.61.1.59
- Guo, S., Lai, C., Wu, C., Cen, G., and Alzheimer's Disease Neuroimaging, I. (2017). Conversion discriminative analysis on mild cognitive impairment using multiple cortical features from MR Images. *Front. Aging Neurosci.* 9:146. doi: 10.3389/fnagi.2017.00146
- Guyon, L., Weston, J., Barnhill, S., and Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Mach. Learn.* 46, 389–422. doi: 10.1023/A:1012487302797
- Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning*. New York, NY: Springer series in statistics.
- Hindmarch, I., Leheld, H., De, J. P., and Erzigkeit, H. (1998). The bayer activities of daily living scale (B-ADL). *Dement. Geriatr. Cogn. Disord.* 9 (Suppl. 2), 20–26. doi: 10.1159/000051195
- Huang, J., and Ling, C. X. (2005). Using, A. U. C., and accuracy in evaluating learning algorithms. *IEEE Trans. Knowl. Data Eng.* 17, 299–310. doi: 10.1109/TKDE.2005.50
- Huang, M., Yang, W., Feng, Q., Chen, W., and Alzheimer's Disease Neuroimaging Initiative. (2017). Longitudinal measurement and hierarchical classification framework for the prediction of Alzheimer's disease. *Sci. Rep.* 7:39880. doi: 10.1038/srep39880
- Jones, E., Oliphant, E., and Peterson, P. (2001). *SciPy: Open Source Scientific Tools for Python*. <http://www.scipy.org/> (Accessed Jan 11, 2017)
- Jung, N.-Y., Seo, S. W., Yoo, H., Yang, J.-J., Park, S., Kim, Y. J., et al. (2016). Classifying anatomical subtypes of subjective memory impairment. *Neurobiol. Aging* 48, 53–60. doi: 10.1016/j.neurobiolaging.2016.08.010
- Kochunov, P., Rogers, W., Mangin, J. F., and Lancaster, J. (2012). A Library of cortical morphology analysis tools to study development, aging and genetics of cerebral cortex. *Neuroinformatics* 10, 81–96. doi: 10.1007/s12021-011-9127-9
- Koepsell, T., and Monsell, S. (2011). Characterizing individuals who revert from mild cognitive impairment to normal or near-normal cognition. *Alzheimer's Dement.* 7:S539. doi: 10.1016/j.jalz.2011.05.1520
- Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., and Baker, C. I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. *Nat. Neurosci.* 12, 535–540. doi: 10.1038/nn.2303
- Lebedev, A. V., Westman, E., Van Westen, G. J., Kramberger, M. G., Lundervold, A., Aarsland, D., et al. (2014). Random forest ensembles for detection and prediction of Alzheimer's disease with a good between-cohort robustness. *Neuroimage Clin.* 6, 115–125. doi: 10.1016/j.nicl.2014.08.023
- Lebedeva, A. K., Westman, E., Borza, T., Beyer, M. K., Engedal, K., Aarsland, D., et al. (2017). MRI-based classification models in prediction of mild cognitive impairment and dementia in late-life depression. *Front. Aging Neurosci.* 9:13. doi: 10.3389/fnagi.2017.00013
- Li, Y., Wang, Y. P., Wu, G. R., Shi, F., Zhou, L. P., Lin, W. L., et al. (2012). Discriminant analysis of longitudinal cortical thickness changes in Alzheimer's disease using dynamic and network features. *Neurobiol. Aging* 33, 427.e15–30. doi: 10.1016/j.neurobiolaging.2010.11.008
- Liu, M., Zhang, J., Yap, P. T., and Shen, D. (2017). View-aligned hypergraph learning for Alzheimer's disease diagnosis with incomplete multi-modality data. *Med. Image Anal.* 36, 123–134. doi: 10.1016/j.media.2016.11.002
- Liu, S. Q., Liu, S. D., Cai, W. D., Che, H. Y., Pujol, S., Kikinis, R., et al. (2015). Multimodal Neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease. *IEEE TBE* 62, 1132–1140. doi: 10.1109/TBME.2014.2372011
- Liu, T., Sachdev, P. S., Lipnicki, D. M., Jiang, J., Geng, G., Zhu, W., et al. (2013). Limited relationships between two-year changes in sulcal morphology and

- other common neuroimaging indices in the elderly. *Neuroimage* 83, 12–17. doi: 10.1016/j.neuroimage.2013.06.058
- Lopez, O. L., Becker, J. T., Chang, Y. F., Sweet, R. A., DeKosky, S. T., Gach, M. H., et al. (2012). Incidence of mild cognitive impairment in the pittsburgh cardiovascular health study-cognition study. *Neurology* 79, 1599–1606. doi: 10.1212/WNL.0b013e31826e25f0
- Maclin, R., and Opitz, D. (1999). Popular ensemble methods: an empirical study. *J. Artif. Intell. Res.* 11, 169–198.
- Macqueen, J. (1967). “Some methods for classification and analysis of multivariate observations,” in *Proceedings of Berkeley Symposium on Mathematical Statistics and Probability* (Berkeley, CA: University of California Press), 281–297.
- Mayo, C. D., Mazerolle, E. L., Ritchie, L., Fisk, J. D., Gawryluk, J. R., and Alzheimer’s Disease Neuroimaging Initiative. (2017). Longitudinal changes in microstructural white matter metrics in Alzheimer’s disease. *Neuroimage Clin.* 13, 330–338. doi: 10.1016/j.nicl.2016.12.012
- McEvoy, L. K., Holland, D., Hagler, D. J. Jr., Fennema-Notestine, C., Brewer, J. B., Dale, A. M., et al. (2011). Mild cognitive impairment: baseline and longitudinal structural MR imaging measures improve predictive prognosis. *Radiology* 259, 834–843. doi: 10.1148/radiol.11101975
- Min, R., Wu, G., Cheng, J., Wang, Q., Shen, D., and Alzheimer’s Disease Neuroimaging Initiative. (2014). Multi-atlas based representations for Alzheimer’s disease diagnosis. *Hum. Brain Mapp.* 35, 5052–5070. doi: 10.1002/hbm.22531
- Misra, C., Fan, Y., and Davatzikos, C. (2009). 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* 44, 1415–1422. doi: 10.1016/j.neuroimage.2008.10.031
- Moradi, E., Pepe, A., Gaser, C., Huttunen, H., Tohka, J., and Alzheimer’s Disease Neuroimaging, I. (2015). Machine learning framework for early MRI-based Alzheimer’s conversion prediction in MCI subjects. *Neuroimage* 104, 398–412. doi: 10.1016/j.neuroimage.2014.10.002
- Mwangi, B., Tian, T. S., and Soares, J. C. (2014). A review of feature reduction techniques in neuroimaging. *Neuroinformatics* 12, 229–244. doi: 10.1007/s12021-013-9204-3
- Pandya, S. Y., Clem, M. A., Silva, L. M., and Woon, F. L. (2016). Does mild cognitive impairment always lead to dementia? A review. *J. Neurol. Sci.* 369, 57–62. doi: 10.1016/j.jns.2016.07.055
- Patenaude, B., Smith, S. M., Kennedy, D. N., and Jenkinson, M. (2011). A Bayesian model of shape and appearance for subcortical brain segmentation. *Neuroimage* 56, 907–922. doi: 10.1016/j.neuroimage.2011.02.046
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Penttilä, J., Paillere-Martinot, M. L., Martinot, J. L., Ringuenet, D., Wessa, M., Houenou, J., et al. (2009). Cortical folding in patients with bipolar disorder or unipolar depression. *J. Psychiatry Neurosci.* 34, 127–135.
- Pereira, F., Mitchell, T., and Botvinick, M. (2009). Machine learning classifiers and fMRI: a tutorial overview. *Neuroimage* 45, S199–S209. doi: 10.1016/j.neuroimage.2008.11.007
- Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. *J. Intern. Med.* 256, 183–194. doi: 10.1111/j.1365-2796.2004.01388.x
- Petersen, R. C. (2010). Does the source of subjects matter? Absolutely! *Neurology* 74, 1754–1755. doi: 10.1212/WNL.0b013e3181e533e7
- Reuter, M., and Fischl, B. (2011). Avoiding asymmetry-induced bias in longitudinal image processing. *Neuroimage* 57, 19–21. doi: 10.1016/j.neuroimage.2011.02.076
- Reuter, M., Schmansky, N. J., Rosas, H. D., and Fischl, B. (2012). Within-subject template estimation for unbiased longitudinal image analysis. *Neuroimage* 61, 1402–1418. doi: 10.1016/j.neuroimage.2012.02.084
- Risacher, S. L., Shen, L., West, J. D., Kim, S., McDonald, B. C., Beckett, L. A., et al. (2010). Longitudinal MRI atrophy biomarkers: relationship to conversion in the ADNI cohort. *Neurobiol. Aging* 31, 1401–1418. doi: 10.1016/j.neurobiolaging.2010.04.029
- Rivière, D., Geffroy, D., Denghien, I., Souedet, N., and Cointepas, Y. (2009). BrainVISA: an extensible software environment for sharing multimodal neuroimaging data and processing tools. *Neuroimage* 47, S163–S163. doi: 10.1016/S1053-8119(09)71720-3
- Rivière, D., Mangin, J. F., Papadopoulos-Orfanos, D., Martinez, J. M., Frouin, V., and Regis, J. (2002). Automatic recognition of cortical sulci of the human brain using a congregation of neural networks. *Med. Image Anal.* 6, 77–92. doi: 10.1016/S1361-8415(02)00052-X
- Roberts, R. O., Knopman, D. S., Mielke, M. M., Cha, R. H., Pankratz, V. S., Christianson, T. J. H., et al. (2014). Higher risk of progression to dementia in mild cognitive impairment cases who revert to normal. *Neurology* 82, 317–325. doi: 10.1212/WNL.0000000000000055
- Rosen, W. G., Mohs, R. C., and Davis, K. L. (1984). A new rating scale for Alzheimer’s disease. *Am. J. Psychiatry* 141, 1356–1364. doi: 10.1176/ajp.141.11.1356
- Rusinek, H., De Santi, S., Frid, D., Tsui, W. H., Tarshish, C. Y., Convit, A., et al. (2003). Regional brain atrophy rate predicts future cognitive decline: 6-year longitudinal MR imaging study of normal aging. *Radiology* 229, 691–696. doi: 10.1148/radiol.2293021299
- Sabuncu, M. R., Desikan, R. S., Sepulcre, J., Yeo, B. T., Liu, H., Schmansky, N. J., et al. (2011). The dynamics of cortical and hippocampal atrophy in Alzheimer disease. *Arch. Neurol.* 68, 1040–1048. doi: 10.1001/archneurol.2011.167
- Sachdev, P. S., Brodaty, H., Reppermund, S., Kochan, N. A., Trollor, J. N., Draper, B., et al. (2010). The Sydney Memory and Ageing Study (MAS): methodology and baseline medical and neuropsychiatric characteristics of an elderly epidemiological non-demented cohort of Australians aged 70–90 years. *Int. Psychogeriatr.* 22, 1248–1264. doi: 10.1017/S1041610210001067
- Schoonheim, M. M., Hulst, H. E., Brandt, R. B., Strik, M., Wink, A. M., Uitdehaag, B. M. J., et al. (2015). Thalamus structure and function determine severity of cognitive impairment in multiple sclerosis. *Neurology* 84, 776–783. doi: 10.1212/WNL.0000000000001285
- Shao, J., Myers, N., Yang, Q., Feng, J., Plant, C., Bohm, C., et al. (2012). Prediction of Alzheimer’s disease using individual structural connectivity networks. *Neurobiol. Aging* 33, 2756–2765. doi: 10.1016/j.neurobiolaging.2012.01.017
- Shi, F., Liu, B., Zhou, Y., Yu, C., and Jiang, T. (2010). Hippocampal volume and asymmetry in mild cognitive impairment and Alzheimer’s disease: meta-analyses of MRI studies. *Hippocampus* 19, 1055–1064. doi: 10.1002/hipo.20573
- Sun, Z. Y., Riviere, D., Poupon, F., Regis, J., and Mangin, J. F. (2007). Automatic inference of sulcus patterns using 3D moment invariants. *Med. Image Comput. Comput. Assist. Interv.* 4791, 515–522. doi: 10.1007/978-3-540-75757-3\_63
- Tabert, M. H., Manly, J. J., Liu, X., Pelton, G. H., Rosenblum, S., Jacobs, M., et al. (2006). Neuropsychological prediction of conversion to Alzheimer disease in patients with mild cognitive impairment. *Arch. Gen. Psychiatry* 63, 916–924. doi: 10.1001/archpsyc.63.8.916
- Tohka, J., Moradi, E., Huttunen, H., and Alzheimer’s Disease Neuroimaging Initiative (2016). Comparison of feature selection techniques in machine learning for anatomical brain MRI in dementia. *Neuroinformatics* 14, 279–296. doi: 10.1007/s12021-015-9292-3
- Trojanowski, J. Q., Vandevertichele, H., Korecka, M., Clark, C. M., Aisen, P. S., Petersen, R. C., et al. (2010). Update on the biomarker core of the Alzheimer’s disease neuroimaging initiative subjects. *Alzheimer’s Dement.* 6, 230–238. doi: 10.1016/j.jalz.2010.03.008
- Trzepak, P. T., Yu, P., Sun, J., Schuh, K., Case, M., Witte, M. M., et al. (2014). Comparison of neuroimaging modalities for the prediction of conversion from mild cognitive impairment to Alzheimer’s dementia. *Neurobiol. Aging* 35, 143–151. doi: 10.1016/j.neurobiolaging.2013.06.018
- Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory*. New York, NY: Springer.
- Walt, S. V. D., Colbert, S. C., and Varoquaux, G. (2011). The NumPy array: a structure for efficient numerical computation. *Comput. Sci. Eng.* 13, 22–30. doi: 10.1109/MCSE.2011.37
- Wee, C.-Y., Yap, P.-T., Li, W., Denny, K., Browndyke, J. N., Potter, G. G., et al. (2011). Enriched white matter connectivity networks for accurate identification of MCI patients. *Neuroimage* 54, 1812–1822. doi: 10.1016/j.neuroimage.2010.10.026
- Wee, C.-Y., Yap, P.-T., Zhang, D., Denny, K., Browndyke, J. N., Potter, G. G., et al. (2012). Identification of MCI individuals using structural and functional connectivity networks. *Neuroimage* 59, 2045–2056. doi: 10.1016/j.neuroimage.2011.10.015
- Wen, W., Sachdev, P. S., Li, J. J., Chen, X. H., and Anstey, K. J. (2009). White Matter Hypertensities in the forties: their prevalence and topography in



- an epidemiological sample aged 44–48. *Hum. Brain Mapp.* 30, 1155–1167. doi: 10.1002/hbm.20586
- Whitwell, J. L., Petersen, R. C., Negash, S., Weigand, S. D., Kantarci, K., Ivnik, R. J., et al. (2007). Patterns of atrophy differ among specific subtypes of mild cognitive impairment. *Arch. Neurol.* 64, 1130–1138. doi: 10.1001/archneur.64.8.1130
- Winblad, B., Palmer, K., Kivipelto, M., Jelic, V., Fratiglioni, L., Wahlund, L. O., et al. (2004). Mild cognitive impairment - beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. *J. Intern. Med.* 256, 240–246. doi: 10.1111/j.1365-2796.2004.01380.x
- Yun, H. J., Kwak, K., and Lee, J. M. (2015). Multimodal discrimination of Alzheimer's Disease based on regional cortical atrophy and hypometabolism. *PLoS ONE* 10:e0129250. doi: 10.1371/journal.pone.0129250
- Zhang, D., Shen, D., and Alzheimer's Disease Neuroimaging Initiative (2012a). Predicting future clinical changes of MCI patients using longitudinal and multimodal biomarkers. *PLoS ONE* 7:e33182. doi: 10.1371/journal.pone.0033182
- Zhang, D., Wang, Y., Zhou, L., Yuan, H., and Shen, D. (2011). Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage* 55, 856–867. doi: 10.1016/j.neuroimage.2011.01.008
- Zhang, H. B., Sachdev, P. S., Wen, W., Kochan, N. A., Crawford, J. D., Brodaty, H., et al. (2012b). Gray matter atrophy patterns of mild cognitive impairment subtypes. *J. Neurol. Sci.* 315, 26–32. doi: 10.1016/j.jns.2011.12.011
- Zhang, J., Liu, M., An, L., Gao, Y., and Shen, D. (2017). Alzheimer's disease diagnosis using landmark-based features from longitudinal structural mr images. *IEEE J. Biomed. Health Informat.* doi: 10.1109/JBHI.2017.2704614. [Epub ahead of print].
- Zhou, J., Liu, J., Narayan, V. A., Ye, J., and Alzheimer's Disease Neuroimaging Initiative (2013). Modeling disease progression via multi-task learning. *Neuroimage* 78, 233–248. doi: 10.1016/j.neuroimage.2013.03.073

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Guan, Liu, Jiang, Tao, Zhang, Niu, Zhu, Wang, Cheng, Kochan, Brodaty, Sachdev and Wen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.