Brain Imaging and Behavior (2021) 15:1728–1738
https://doi.org/10.1007/s11682-020-00366-8

ORIGINAL RESEARCH

Predicting Alzheimer's conversion in mild cognitive impairment patients using longitudinal neuroimaging and clinical markers

Carlos Platero[1] · M. Carmen Tobar[1] · for the Alzheimer's Disease Neuroimaging Initiative

Published online: 10 November 2020

Abstract

Patients with mild cognitive impairment (MCI) have a high risk for conversion to Alzheimer's disease (AD). Early diagnose of AD in MCI subjects could help to slow or halt the disease progression. Selecting a set of relevant markers from multimodal data to predict conversion from MCI to probable AD has become a challenging task. The aim of this paper is to quantify the impact of longitudinal predictive models with single- or multisource data for predicting MCI-to-AD conversion and identifying a very small subset of features that are highly predictive of conversion. We developed predictive models of MCI-to-AD progression that combine magnetic resonance imaging (MRI)-based markers (cortical thickness and volume of subcortical structures) with neuropsychological tests. These models were built with longitudinal data and validated using baseline values. By using a linear mixed effects approach, we modeled the longitudinal trajectories of the markers. A set of longitudinal features potentially discriminating between MCI subjects who convert to dementia and those who remain stable over a period of 3 years was obtained. Classifier were trained using the marginal longitudinal trajectory residues from the selected features. Our best models predicted conversion with 77% accuracy at baseline (AUC = 0.855, 84% sensitivity, 70% specificity). As more visits were available, longitudinal predictive models improved their predictions with 84% accuracy (AUC = 0.912, 83% sensitivity, 84% specificity). The proposed approach was developed, trained and evaluated using the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset with a total of 2491 visits from 610 subjects.

Keywords Alzheimer's disease · MRI · Longitudinal analysis

Introduction

Alzheimer's disease (AD) is the most common neurodegenerative disorder in the elderly, characterized by pathological changes in the brain that begin 10 ∼ 15 years prior to the onset of clinical symptoms (Jack et al. 2010). Mild cognitive impairment (MCI), the earliest clinically detectable phase of the trajectory toward dementia and AD, affects 15% to 20% of people older than 65 years (Markesbery 2010; Roberts and Knopman 2013). Approximately 10-15% of MCI patients will develop into dementia annually (Petersen et al. 2009). Predicting the conversion from MCI into probable AD could help to slow or halt disease progression, being one of the most challenging tasks in the field of AD (Jack 2012).

An important task to improve the diagnosis of AD is the biomarker selection from several data modalities such as clinical, imaging, genetic and fluid data (Da et al. 2014; Moradi et al. 2015; Korolev et al. 2016; Gavidia-Bovadilla et al. 2017). MRI-derived measurements of the brain have become useful markers in the diagnosis of AD and MCI (Cuingnet et al. 2011; Rathore et al. 2017). Markers based on MRI have a minimal cost impact because MRI scanning is often part of the clinical assessment standard for patients with MCI (Albert et al. 2011). The volume of the hippocampus is the most studied and used structural MRI marker of AD up to date (Cuingnet et al. 2011). Markers applied in different parts of the brain are expected to be more sensitive to different stages of the AD (Li et al. 2012; Eskildsen et al. 2015; Moradi et al.

Membership of the Alzheimer's Disease Neuroimaging Initiative (ADNI) is provided in the Acknowledgments.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11682-020-00366-8) contains supplementary material, which is available to authorized users.

✉ Carlos Platero
carlos.platero@upm.es

[1] Health Science Technology Group, Universidad Politécnica de Madrid, Ronda de Valencia 3, 28012, Madrid, Spain

2015). Accordingly, the combination of cortical thickness (CT) and volume of subcortical structures has been used in several studies (Korolev et al. 2016; Sørensen et al. 2017). Recently, CT has been proposed as a more stable parameter for AD diagnosis than volume/density measures (Dickerson et al. 2009; Desikan et al. 2009; Li et al. 2012; Park et al. 2013; Bernal-Rusiel et al. 2013; Eskildsen et al. 2013; Pettigrew et al. 2016).

AD is a progressive neurodegenerative disorder, and longitudinal data could improve the predictive power (Gavidia-Bovadilla et al. 2017; Minhas et al. 2017). Over time, MRI-derived longitudinal measurements have been shown to correlate with the progression of AD (Wolz et al. 2010; Aubert-Broche et al. 2013; Iglesias et al. 2016; Platero et al. 2018). Tracking longitudinal brain atrophy in these neurodegenerative conditions has only recently become feasible, with longitudinal methods allowing the analysis of univariate and mass-univariate neuroimaging measures based on linear mixed effects (LME) modeling (Bernal-Rusiel et al. 2013; Bernal-Rusiel et al. 2013). In this way, we combined MRI-based markers and standard neuropsychological measures (NMs) for predicting MCI-to-dementia progression using longitudinal data and validated with baseline scores from the subjects. Our goals were to (1) improve the performance of the predictive models of MCI-to-AD progression and (2) find a group of interpretable features that are highly predictive of conversion.

Materials

The Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset was selected to evaluate the behavior of the longitudinal classification framework for predicting MCI-to-AD conversion, where subjects have different numbers of visits (Wyman et al. 2013; Weiner and Veitch 2015). MRI-based markers and NMs used in this study correspond to measures of neurodegeneration at 3 years follow-up. We included an unbalanced longitudinal data of 610 subjects at multiple time points: baseline and 6, 12, 18, 24 and 36 months.

Since our main objective was to predict the conversion of MCI to dementia, two different strategies could be used for the construction of predictive models. The first one was to use a population of subjects probable AD and normal controls (NC). Subsequently, given the feature selection from NC and AD data, the population of stable MCI (sMCI) and progressive MCI (pMCI) subjects was used to validate the predictive models (Da et al. 2014; Moradi et al. 2015; Eskildsen et al. 2015; Beheshti et al. 2017). The second strategy was to exclusively use the population of sMCI and pMCI subjects with a nested cross-validation procedure to estimate the feature selection and performances of the predictive models (Korolev et al. 2016). An overview of the subject groups is given in Table 1. Table 2 reports the time points (Baseline, Month 6, Month 12, Month 24 and Month 36) that were available for the selected subjects.

Method

From a two-group comparison approach, which is applied to the sMCI and pMCI patients, two types of predictive models were built and examined. The first type of predictive models was constructed using a single source, MRI data or NMs. The second type of models used multisource data, i.e., predictive models built with MRI-based features and NMs. For each proposed model, sensitivity, specificity and accuracy scores of the classifiers were computed (Cuingnet et al. 2011). Additionally, receiver operating characteristic (ROC) curves were also calculated. The discriminant value of the corresponding ROC curve was estimated using the area under the curve (AUC). We also examined the effects of age, sex, educational level and APOE genotype as covariates in the models in terms of predictive performance.

We used LME modeling to account the between-subject and within-subject sources of variation (Bernal-Rusiel et al. 2013; Bernal-Rusiel et al. 2013).The LME model was built with a intercept and slope as random effects to be included in the longitudinal trajectory:

$$\begin{aligned} y_{ij} &= (\beta_1 + \beta_2 \cdot Group_i + \beta_3 \cdot Age_i + \beta_4 \cdot Education_i \\ &\quad + \beta_5 \cdot Sex_i + \beta_6 \varepsilon 4_i + b_{ri}) \\ &\quad + (\beta_7 + \beta_8 \cdot Group_i + \beta_9 \cdot \varepsilon 4_i + b_{si})\, t_{ij} + e_{ij}, \end{aligned} \quad (1)$$

where $j = 1, ..., n_i$ indexes the time points with n_i indicating the number of scans for subject i, y_{ij} is the j measure of a feature from subject i, t_{ij} is the scan time from baseline (in years), and $\beta_r = [\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6]^T$ and $\beta_s = [\beta_7, \beta_8, \beta_9]^T$ are the intercept and slope, respectively. The components of $b_i = [b_{si}, b_{ri}]^T$ is a vector of the random effects, and e_i is a vector of measurement errors. The boolean variable $Group_i$ is true if the i-subject progresses to AD and false if the subject remains stable or is control. The influence of sociodemographic characteristics was collected with the effect of age at baseline (Age_i), sex (Sex_i) and years of education ($Education_i$) (Jiang et al. 2014; Liu et al. 2012). Apolipoprotein E (APOE) genotype status is the most prevalent genetic risk factor for AD (Saykin et al. 2010). APOE genotype was assessed, and patients were characterized as $\varepsilon 4$ carriers or $\varepsilon 4$ noncarriers. The interaction between APOE genotype status and time was also included based on the evidence that $\varepsilon 4$ accelerates atrophy during the prodromal phases of AD (Bernal-Rusiel et al. 2013).

Table 1 Demographic and clinical details of the subset of ADNI database used in this study

Type (N. subjects)	NC (113)	sMCI (215)	pMCI (206)	AD (76)	F	p
Sex male (%)	72 (64%)	146 (68%)	121 (59%)	46 (61%)		0.144
Baseline age	75.6 (5.2)	74.9 (7.5)	75.0 (6.9)	73.8 (6.9)	1.02	0.282
MMSE (Baseline)	29.2 (1.0)	27.6 (1.7)^{a}	26.6 (1.8)^{a,b}	23.4 (1.8)^{a,b,c}	209.84	< 0.001
MMSE (Month 6)	29.1 (1.0)	27.6 (2.2)^{a}	25.3 (2.9)^{a,b}	22.3 (3.1)^{a,b,c}	141.76	< 0.001
MMSE (Month 12)	28.8 (3.0)	27.5 (3.9)^{a}	24.9 (2.9)^{a,b}	20.6 (4.5)^{a,b,c}	94.94	< 0.001
MMSE (Month 24)	29.2 (0.9)	27.1 (4.7)^{a}	22.7 (4.5)^{a,b}	18.1 (6.9)^{a,b,c}	81.7	< 0.001
MMSE (Month 36)	–	26.2 (6.3)	21.7 (3.7)^{b}	–	11.8	< 0.001

Data as represented as mean and standard deviation (SD) unless specific otherwise. ANOVA with Bonferroni post hoc test is used for baseline age and neuropsychological score, except for sex where the chi-square test is used. Statistical significance is considered with $p - value < 0.01$. [a] Significant compared to normal control (NC). [b] Significant compared to sMCI. [c] Significant compared to pMCI. NC= Normal control; sMCI= Stable Mild cognitive impairment; pMCI= Progressive Mild cognitive impairment; AD= Alzheimer disease; MMSE=Mini-Mental State Examination

For a feature, the difference between the longitudinal trajectory of an i-subject and the LME model will be described by the random vectors b_i and e_i, which follow mean zero-Gaussian multivariate distributions, indicating a population-averaged mean of $E(Y_i) = X_i\beta$ (Bernal-Rusiel et al. 2013). Therefore, the longitudinal trajectory residue is defined as follows:

$$l_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\left(y_{ij} - (X_i)_j\beta\right), \tag{2}$$

where $(X_i)_j$ is the j-row vector of the design matrix and the boolean variable of the clinical group is activated $(Group_i = 1)$, i.e., the effects of the AD progression compared to control or stable MCI are considered. The l_i-samples belonging to the AD or pMCI group follow a normal distribution of zero mean and of variance determined by b_i and e_i. In contrast, l_i-samples not belonging to the AD or pMCI group will also follow normal a distribution with the same variance but with a bias in relation to zero-mean, which is related to the fixed effects of the clinical groups. The random variable l_i is used to train and classify the features by linear discriminant analysis (LDA). All longitudinal trajectory residues of the selected features were assumed to be independent. These marginal residues were used as data for training and testing of the LDA classifiers. Note that these classifiers do not require the adjustment of any parameter, as is traditionally done at this stage (Moradi et al. 2015; Korolev et al. 2016). The marginal residues of the longitudinal trajectories of the markers as inputs to the LDA were experimentally validated (see Supplementary Materials S.4).

A nested cross-validation (CV) procedure was used to avoid model overfitting and optimistically biased estimates of model performance (Korolev et al. 2016). The procedure consisted of two nested CV loops: an inner loop, designed to select the optimal feature set for the proposed models, and an outer loop, designed to obtain an unbiased estimate of model performance. In the outer CV loop, the data were partitioned into the model and test data (see Fig. 1). In the inner CV loop, the model data were again partitioned into the training and validation data. Random subsets of training data were subjected to the minimal-redundancy-maximal-relevance (mRMR) algorithm (Peng et al. 2005). For each sample, subsets of features with different dimensions were suggested. This sequence (i.e., selecting random subsets

Table 2 Number and timing of scans per time point by clinical group

Type (N. subjects)	NC	sMCI	pMCI	AD	Time from baseline (in month)
Baseline	113	215	206	76	0
Month 6	111	205	197	76	6.9(0.8)
Month 12	108	200	191	76	12.9(0.8)
Month 18	0	92	88	0	19.0(0.9)
Month 24	88	149	135	46	25.0(0.8)
Month 36	0	66	54	0	36.9(1.1)
Total	420	865	932	274	

NC= Normal control; sMCI= Stable Mild cognitive impairment; pMCI= Progressive Mild cognitive impairment; AD= Alzheimer disease

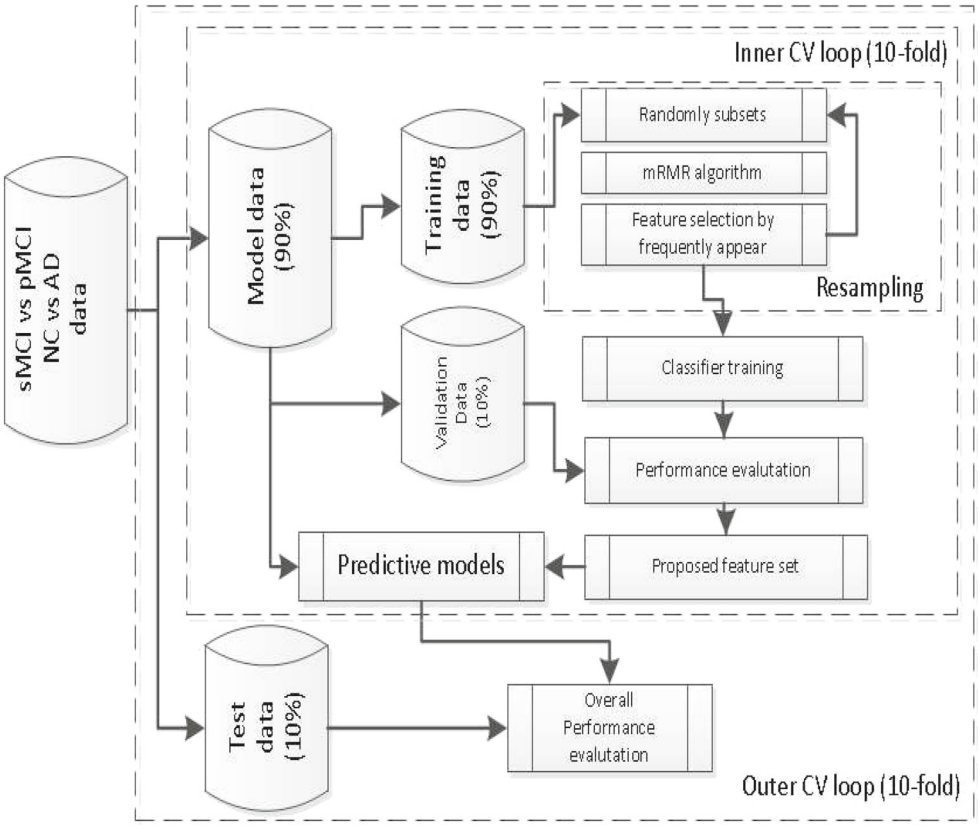

Fig. 1 Nested 10-fold cross-validation procedure for model development and evaluation using MCI subjects. In the case of adding the NC and AD populations, these patients were used only in the feature selection, i.e. their data were used in the inner CV loop

of training, applying mRMR and proposing combination features) was repeated several times. For each dimension, the combinations of features that most frequently appear were selected. Then, predictive models were trained using only the training data with the candidate feature set. The best combinations of features, which produced the maximal classification accuracy, were selected to build the final predictive models. These models were constructed by training a classifier using longitudinal data. An unbiased estimate of model performances was obtained by evaluating the final models on the withheld test data, which were not used during feature selection, model selection, or final model construction. Both the outer and inner CV loops used a 10-fold CV design. For better replicability, the nested 10-fold CV procedure was repeated with different partitions of the data, generating multiple performance estimate values. A MATLAB implementation of our method is available at https://www.nitrc.org/projects/predict_mci2ad/.

Results

A total of 2491 scans from 610 subjects were processed. For each each subject, subcortical volumes, ROI-based and cluster-based CT measurements were obtained using the longitudinal pipeline (see Supplementary Materials S.1). To validate the consistency of the applied longitudinal pipeline, a quality control was used between the scans and segmentations of each subject (see Supplementary Materials S.2).

Cortical thickness analysis

By comparing the spatial patterns of cortical thinning, it was found that the spatial patterns show annual rates of atrophy different in the clinical groups of the progression of AD (Dickerson et al. 2009; Li et al. 2012; Pettigrew et al. 2016). Therefore, we constrained our analyses to the association between the group time interaction and cortical thickness.

The cohort used in this study were divided into two independent populations. The first population was used to build statistical maps of annual atrophy rates between NC and AD or sMCI and pMCI. The second patient population was used to obtain the predictive models between the two clinical groups of MCI. This strategy was used to avoid contamination between the developments of the statistical significance maps and predictive models. Using only MCI patients, the first population consisted of 50/50 subjects of sMCI and pMCI with a total of 467 scans, which were available from the ST-LME toolbox (Bernal-Rusiel et al. 2013). The second population consisted on 165/156 sMCI

and pMCI patients with 1330 visits, which were used for building and evaluating the predictive models. In contrast, we used 40/30 NC and AD subjects with 278 visits to obtain the statistical maps and 73/46 NC and AD subjects with 416 visits in order to develop the predictive models from the feature selection stage.

To generate statistical significance maps for characterizing longitudinal thinning differences between two clinical groups, an empirical strategy was used (Thirion et al. 2007). Randomly subsets of subjects belonging to the two clinical groups were conducted to the ST-LME method. A two-stage adaptive false discovery rate (FDR) procedure was employed to control multiple comparisons (Benjamini et al. 2006). Before calculating the statistical maps of cortical thinning rates between two clinical groups, we proceeded to test the control of type I and II errors of the ST-LME method (see Supplementary Materials S.3). A separate hypothesis test at each vertex was conducted in the ST-LME approach (see Fig. 2a). A binary map was derived by thresholding the values shown in the statistical map with an FDR correction. Within this thresholded map, only clusters that reached a threshold region size of a set of contiguous vertices were retained ($> 100mm^2$) (see Fig. 2b). Cortical features were determined as the mean CT for each selected cortical cluster.

Performance of the predictive models of MCI-to-AD progression

The feature extraction stage obtained 40 ROI-based MRI markers, 11 NMs and approximately a dozen clusters for each hemisphere. Once the features were extracted, the nested k-fold CV procedure was performed to select the best subsets of features and to build the predictive models with their subsequent evaluations. The nested 10-fold CV procedure was repeated 50 times with different partitions of the data. In the feature selection stage, feature subsets of different dimensions were defined using the mRMR algorithm. We used the mutual information difference metric, and the features were normalized to zero-mean and unit-variance in the mRMR algorithm (Peng et al. 2005).

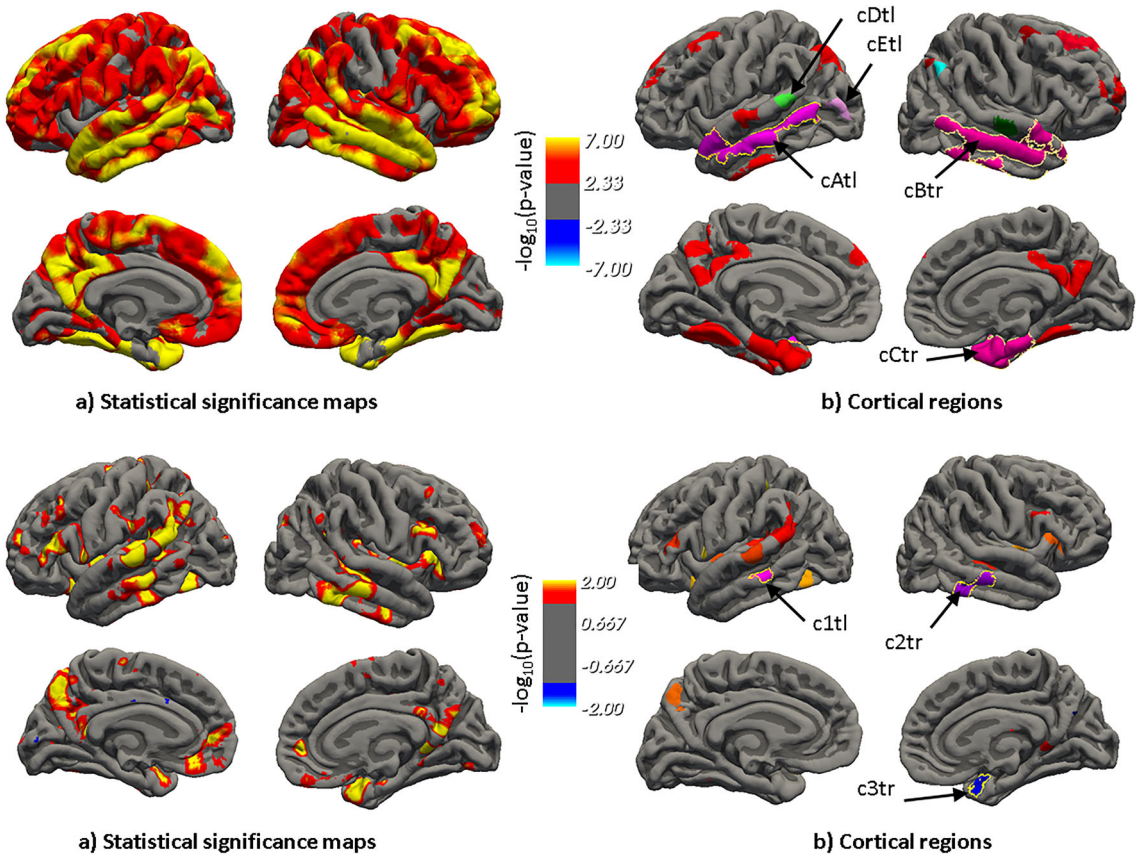

Fig. 2 **a** Statistical significance maps ($-\log_{10}(p - value)$) comparing longitudinal cortical thinning rates between NC and AD patients (first row) and sMCI and pMCI patients (second row) visualized on the pial surface of the FreeSurfer template (fsaverage). **b** Cortical cluster exhibiting a statistically significant difference in longitudinal thinning between two clinical groups (NC-AD and sMCI-pMCI). These maps were derived by thresholding the values shown with a FDR correction at $q < 10^{-7}$ (NC and AD) and $q < 0.01$ (sMCI and pMCI). NC= Normal control; sMCI=Stable Mild cognitive impairment; pMCI=Progressive Mild cognitive impairment; AD=Alzheimer disease

For each inner loop, the resampling method searched for the 10-feature subset of each dimension that appears most frequently in mRMR. Of these, the 3 that best offered their performance in classification accuracy terms by the evaluation of the corresponding models were selected. Therefore, for each outer iteration, 30 feature subsets for each dimension were evaluated. In total, there were 15,000 evaluations of selected subsets for each dimension. Predictive models with higher AUC values and balanced between sensitivity and specificity were selected. Table 3 summarizes the performance of the predictive models with baseline data. Classification results are presented with predictive models constructed with single- or multisource data. For each source of data used, we present the best predictive models identified with their markers, well trained exclusively with MCI population or having used the NC and AD population to select the features. Several predictive models instead of one for each type of training are presented due to these proposed models exhibiting combinations of markers with similar performance. The LME models were built using only age at baseline as covariate. Table 3 shows the scores of the first combination of markers for each type of training.

The performances of the predictive models using MRI data, which were trained with NC and AD and with sMCI and pMCI to define the feature selection, were similar and almost coincident for the selected markers. MRI-based markers included left hippocampal volume and CT in medial temporal lobe (MTL) and left inferior parietal lobule. Markers from CT in massive measures on hemispheres were added to the set of ROI-based MRI features. We observed that the cluster-based CT markers were first selected with respect to the ROI-based CT markers by the mRMR algorithm. In all our experiments using only MRI data, better classification accuracies were obtained by combining CT measures with subcortical volumes normalized by intracranial volume, as in other publications (Westman et al. 2013). When predictive models were built with only cognitive measures, ADAS13 and FAQ were selected as the best combination of features using NC and AD data. In contrast, using the sMCI and pMCI data, ADAS13, FAQ and RAVLT immediate were selected, yielding an improvement in the classification results. The best results of the predictive models were obtained when combining the MRI-based markers and NMs, especially when the feature selection was performed with the data from the sMCI and pMCI populations. The models were basically built by combining ADAS13, FAQ and RAVLT immediate with hippocampal volumetry and thinning measures in the temporal lobe. We also observed that the models trained exclusively with sMCI and pMCI populations outperformed the models built with NC and AD subjects due to better selection in the cognitive measures.

These baseline classification results exclusively used age as a covariate in the LME models. Once we observed that the best results were obtained using combinations of MRI-based markers and NMs, more covariates were added to the LME models. Table 4 shows that the best results were achieved by adding age, sex and years of education as covariates. The selected feature subsets did not change practically when new covariates were added to the LME models.

The last experiment was to analyze the performance of predictive models when data were available from new patient visits. Table 5 shows the scores of the predictive models previously proposed (see Table 4) with age, sex and years of education as covariates in LME models applied to longitudinal data. The predictive models improved AD conversion prediction as more patient data were available over time. The scores improved with the increase in patient visits, and the balance between sensitivity and specificity was also improved.

Discussion

The analysis of 610 participants with 2491 visits provided a large number of subjects for the training and testing datasets. The NM-based predictive models outperformed the MRI models on accuracy (NM models: AUC = 0.826, ACC = 75%, SEN = 86%, SPE = 65%, MRI models: AUC = 0.778, ACC = 72%, SEN = 77%, SPE = 67%). The multisource model outperformed both single-source models (AUC = 0.855, ACC = 77%, SEN = 84%, SPE = 70%). All these scores were at baseline. As more visits were available, longitudinal predictive models improved their predictions (AUC = 0.912, ACC = 84%, SEN = 83%, SPE = 84 %). The specificity, accuracy and AUC positively increased their scores with the time of follow-up, maintaining the sensitivity scores. There is also a better balance between sensitivity and specificity over time.

The feature vector dimensions were very low in comparison to other studies. Low dimensions in predictive models make them more robust, with more generalization capacity and lower risk of overfitting. The features most frequently selected in the NM model included ADAS13, FAQ and RAVLT immediate. In the MRI model, the most frequently selected features included hippocampal volume and CT measures for several temporoparietal brain regions, with a preference for the left hemisphere. The selection of hippocampus, MTL, and inferior parietal cortex as predictors of MCI-to-dementia progression is consistent with the known pattern of grey matter atrophy associated with incipient AD (Thompson et al. 2003), and there is also evidence that AD-related atrophy occurs at a faster rate in the left hemisphere (Thompson et al. 2003). Similar findings have been

Table 3 Baseline scores for predicting MCI-to-AD conversion

Data	AUC	ACC (%)	SEN (%)	SPE (%)	Features
MRI(ROI) NC-AD	0.758 (0.754 0.763)	72 (71.7 72.5)	77.7 (77.1 78.2)	66.8 (66.2 67.4)	
					Hvl,IPtl,Etr
					Hvl,IPtl,AVtb,Etr
					Hvl,IPtl,Etr
MRI(ROI) s-pMCI	0.763 (0.759 0.767)	71.8 (71.5 72.2)	77.0 (76.4 77.5)	67.0 (66.4 67.6)	
					Hvl,IPtl,Etr
					Hvl,IPtb
					Hvb,IPtl
MRI(all) NC-AD	0.778 (0.774 0.782)	71.4 (71.1 71.8)	76.3 (75.7 76.8)	66.9 (66.2 67.5)	
					Hvl,cBtr,cAtl
					Hvl,Etb,IPtb,cCtr
					Hvl,AVtb,cBtr
MRI(all) s-pMCI	0.761 (0.757 0.765)	67.8 (67.4 68.2)	72.6 (72.1 73.2)	63.4 (62.7 64.0)	
					Hvl,Pvl,c2tr
					Hvl,c2tr
					Hvl,MTtb
NM NC-AD	0.816 (0.814 0.821)	73.0 (72.6 73.4)	83.8 (83.2 84.5)	63.0 (62.2 63.7)	
					A13,FAQ
					A13,FAQ,MMSE
NM s-pMCI	0.826 (0.822 0.830)	75.4 (75.1 75.8)	86.2 (85.8 86.7)	65.3 (64.6 65.9)	
					A13,FAQ,RTim
					A13,FAQ,RTle
MRI(ROI)+NM NC-AD	0.826 (0.822 0.830)	73.3 (72.9 73.7)	83.7 (83.1 84.3)	63.6 (63.0 64.3)	
					Hvl,ITtl,A11,AQ4,FAQ,MMSE
					Hvl,MTtb,A11,AQ4,FAQ,MMSE
					ITtl,AQ4,FAQ
MRI(ROI)+NM s-pMCI	0.855 (0.851 0.858)	76.9 (76.5 77.3)	84.4 (83.7 85.0)	69.9 (69.1 70.6)	
					Hvl,Pvl,MTtb,A13,FAQ,RTim,CDR
					Hvl,Pvl,MTtb,A13,FAQ,RTim,MMSE
					Pvl,MTtb,A13,FAQ,RTim,CDR
MRI(all)+NM NC-AD	0.833 (0.830 0.836)	73.0 (72.6 73.3)	83.1 (82.5 83.7)	63.4 (62.8 64.1)	
					Hvl,MTtb,A11,AQ4,FAQ,MMSE
					Hvl,cEtl,A11,AQ4,FAQ,MMSE
					Hvl,cDtl,A11,AQ4,FAQ,MMSE
MRI(all)+NM s-pMCI	0.846 (0.843 0.849)	75.2 (74.8 75.6)	82.2 (81.5 82.8)	68.7 (67.9 69.6)	
					Pvl,c2tr,A13,FAQ,RTim,CDR
					Hvl,Pvl,c2tr,A13,FAQ,RTim
					Hvl,c1tl,c3tr,A13,FAQ,RTim

Each row shows the classification results of the predictive models according to the type of data used (MRI, NM or MRI + NM). Numbers within parentheses are the 95% confidence interval. The notation of the MRI-based features is given by an uppercase acronym of the brain structure followed by two lowercase letters. The first lowercase letter indicates a measure of volumetry (v) or cortical thickness(t). The second lowercase letter indicates that the measurement is from the left hemisphere (l), right (r) or bilateral (b). The cluster-based CT measurements follow the same notation. Their labels are shown in the Fig. 2. H=Hippocampus; P=Pallidum; MT=Middle Temporal; IT=Inferior Temporal; IP=Inferior Parietal; E=Entorhinal; AV= AD vulnerable (Pettigrew et al. 2016); A11=ADAS-Cog 11; A13=ADAS-Cog 13; AQ4=ADAS Q4; RTim=RAVLT immediate; RTle=RAVLT learning; NC= Normal control; sMCI=Stable Mild cognitive impairment; pMCI=Progressive Mild cognitive impairment; AD=Alzheimer disease; AUC=Area under curve; ACC=accuracy; SEN=sensitivity; SPE=specificity

Table 4 Baseline scores for predicting MCI-to-AD conversion with different covariates using MRI-markers and NMs

AUC	ACC (%)	SEN (%)	SPE (%)	Features
0.855 (0.851 0.858)	76.9 (76.5 77.3)	84.4 (83.7 85.0)	69.9 (69.1 70.6)	
				$Hvl, Pvl, MTtb, A13, FAQ, RTim, CDR$
				$Hvl, Pvl, MTtb, A13, FAQ, RTim, MMSE$
				$Pvl, MTtb, A13, FAQ, RTim, CDR$
0.848 (0.845 0.851)	76.5 (76.1 76.9)	82.9 (82.2 83.6)	70.5 (69.7 71.2)	
				$Hvl, Pvl, MTtb, A13, FAQ, RTim, CDR$
				$Hvl, Pvl, MTtb, A13, FAQ, RTim, MMSE$
				$Pvl, MTtb, A13, FAQ, RTim, CDR$
0.858 (0.855 0.862)	77.3 (77.0 77.7)	84.7 (84.0 85.3)	70.5 (69.8 71.2)	
				$Hvl, Plv, MTtb, A13, FAQ, RTim, CDR$
				$Hvl, Pvl, MTtb, A13, FAQ, RTim$
				$Pvl, MTtb, A13, FAQ, RTim, CDR$
0.851 (0.848 0.855)	77.0 (76.6 77.4)	83.2 (82.5 83.9)	71.2 (70.5 71.9)	
				$Hvl, Pvl, MTtb, A13, FAQ, RTim, CDR$
				$Hvl, Pvl, MTtb, A13, FAQ, RTim$

The first row using only age as a covariate, in the second row age and APOE $\varepsilon 4$, in the third row age, sex and education and in the fourth row, age, sex, education and APOE $\varepsilon 4$. Numbers within parentheses are the 95% confidence interval. The notation of the MRI-based features is given by an uppercase acronym of the brain structure followed by two lowercase letters. The first lowercase letter indicates a measure of volumetry (v) or cortical thickness (t). The second lowercase letter indicates that the measurement is from the left hemisphere (l), right (r) or bilateral (b). H=Hippocampus; P=Pallidum; MT=Middle Temporal; A11=ADAS-Cog 11; A13=ADAS-Cog 13; RTim=RAVLT immediate; AUC=Area under curve; ACC=accuracy; SEN=sensitivity; SPE=specificity

reported in other studies (Korolev et al. 2016; Sørensen et al. 2017). Using multisource predictive models, the best classification results were obtained when constructing LME models that added age, sex and years of education as covariates. APOE genotype was not selected as covariates in LME approach. This result was consistent with APOE $\varepsilon 4$ being a risk factor for AD, however its value for individual patient predictions was limited (Da et al. 2014).

CT measures were extracted from the ROIs and clusters. The statistical maps were in strong agreement with the literature (Bernal-Rusiel et al. 2013; Landin-Romero et al. 2017). Specifically, our longitudinal analyses revealed that using NC and AD data for building statistical maps, the proposed predictive models selected cluster-based CT measurements in middle and superior temporal regions and inferior parietal regions. When the predictive models were constructed exclusively with sMCI and pMCI patients, the clusters were bilaterally in the MTL. These findings are consistent with neuropathological studies showing that neurofibrillary tangles initially accumulate in MTL regions

Table 5 Scores for predicting MCI-to-AD conversion with different visits

Time	AUC	ACC (%)	SEN (%)	SPE (%)
Baseline	0.858(0.855 0.862)	77.3(77.0 77.7)	84.7(84.0 85.3)	70.5(69.8 71.2)
Month 6	0.879(0.876 0.882)	79.2(78.9 79.6)	84.8(84.3 85.4)	74.0(73.4 74.6)
Month 12	0.892(0.889 0.894)	80.9(80.6 81.2)	83.8(83.2 84.3)	78.4(77.8 79.0)
Month 24	0.905(0.902 0.908)	82.0(81.6 82.3)	78.6(78.0 79.3)	84.6(84.0 85.2)
Month 36	0.928(0.923 0.933)	85.9(85.3 86.5)	79.3(77.9 80.7)	89.8(89.0 90.6)
longitudinal	0.912(0.909 0.914)	83.7(83.4 84.0)	83.4(82.9 84.0)	84.0(83.5 84.5)

Numbers within parentheses are the 95% confidence interval. The longitudinal results correspond to the classification scores with the unbalanced data of test subjects. The predictive models used in this experiment correspond to those indicated in Table 4 with age, sex and years of education as covariates in LME models. AUC=Area under curve; ACC=accuracy; SEN=sensitivity; SPE=specificity

Table 6 Comparison of approaches for predicting MCI-to-AD conversion

Approach	Follow-up	Subject number sMCI/pMCI	Data	AUC	ACC (%)	SEN (%)	SPE (%)	Number of features
Gavidia-Bovadilla et al. (2017)	36		MRI		70.3	64.3	75.2	
Beheshti et al. (2017)	36	65/71	MRI	0.751	75.0	76.9	73.2	
Eskildsen et al. (2015)	36	227/161	MRI	0.763	71.9	69.6	73.6	5
Moradi et al. (2015)	36	164/100	MRI	0.766	74.7	88.8	51.6	309
Minhas et al. (2017)	24	65/54	MRI	0.811	77.5	53.9	89.2	38
Present study	36	215/206	MRI	0.763	71.8	77.0	67.0	3
				0.795	74.1	76.8	71.5	
Ferreira et al. (2017)	36	160/122	NM	0.870	81.6	85.0	79.0	7
Minhas et al. (2017)	36	65/37	NM	0.881	83.3	85.6	81.5	8
Present study	36	215/206	NM	0.826	75.4	86.2	65.3	3
				0.896	80.6	84.4	80.6	
Korolev et al. (2016)	36	120/139	MRI,NM	0.87	79.9	83.4	76.4	10
Moradi et al. (2015)	36	164/100	MRI,NM	0.902	82.0	87.0	74.0	309
Gavidia-Bovadilla et al. (2017)	36		MRI,NM		76.7	70.8	81.6	
Minhas et al. (2017)	24	65/54	MRI,NM	0.889	84.3	70.4	92.3	45
Present study	36	215/206	MRI,NM	0.858	77.3	84.7	70.5	6
				0.912	83.7	83.4	84.0	

Our classification results are given at baseline as well as the scores from unbalanced longitudinal trajectories. Our predictive models from single source data were built using LME models with only age at baseline as covariate and models with multi-source data used age, sex and years of education as covariates. sMCI=Stable Mild cognitive impairment; pMCI=Progressive Mild cognitive impairment; AUC=Area under curve; ACC=accuracy; SEN=sensitivity; SPE=specificity

before spreading to other regions, including adjacent temporal and parietal areas (Braak and Braak 1991; Pettigrew et al. 2016). We also observed that models exclusively built with the sMCI and pMCI population yielded better benefits than those trained with feature selection based on the populations of NC and AD, especially in the choice of the neuropsychological measures.

In recent years, many studies have been published in the field of MCI-to-AD prediction. Reviews of these approaches can be found in Falahati et al. (2014) and Rathore et al. (2017). Table 6 shows that our best prediction models, with single- or multisource data, performed very favorably compared with recently published models. For better compatibility with the present study, we limit this comparison to studies that used data from the ADNI dataset to predict MCI-to-AD progression from 24 to 36 months of follow-up.

Limitations

Despite promising results, there are several limitations of our study. The data used here correspond to subjects who meet the inclusion and exclusion criteria established by ADNI. The choice of cohort may affect the results of the predictive models. It would remain as future work to apply the proposed approach for developing new predictive models using other public cohorts, like OASIS-3 (LaMontagne et al. 2018).

The modeling of markers was implemented using the LME approach. A linear function to model the dynamic changes of structural MRI-based markers is a well-accepted practice (Guerrero et al. 2016). Experimental evidence suggests that linear models are not enough to describe cognitive decline measurements in AD progression (see Supplementary Materials S.5). A future line of work would be to improve the modeling of marker trajectories.

Conclusions

Some conclusions of this study were the following: 1) Identifying most robust predictors of conversion using feature subsets that appear most frequently in mRMR and better performance in classification; 2) Classification of MCI patients into sMCI and pMCIs using the marginal longitudinal trajectory residues from the selected features; 3) The proposed predictive models were built with only 3-6 features highly stable under cross validation; these markers were consistent with Braak stages and previously reported in MCI-to-AD conversion and are easy to transfer to new cohorts and clinical practice; 4) Multisource data for predicting MCI-to-AD conversion

deliver a more accurate estimate than single-source data; 5) The method uses relatively common clinical tests such as MRI and neuropsychological tests, as opposed to methods that rely on more expensive or invasive tests such as PET-based, CSF-based and genotype-based markers; 6) It can be observed that longitudinal NM and MRI markers embedded in our proposed longitudinal framework outperform other classification methods; 7) These reliable NM and MRI markers of AD progression can offer potential for monitoring treatment outcome in future drug trials; and 8) The proposed approach was developed, trained and evaluated using the ADNI dataset.

Acknowledgments Data used in preparation of this article were obtained from the ADNI database (adni.loni.usc.edu). 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: http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Funding This research did not receive any specific grant.

Compliance with Ethical Standards

Conflict of interests The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Albert, M.S., DeKosky, S.T., Dickson, D., Dubois, B., Feldman, H.H., Fox, N.C., et al. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, *7*(3), 270–279.

Aubert-Broche, B., Fonov, V., García-Lorenzo, D., Mouiha, A., Guizard, N., Coupé, P., et al. (2013). A new method for structural volume analysis of longitudinal brain MRI data and its application in studying the growth trajectories of anatomical brain structures in childhood. *Neuroimage*, *82*, 393–402.

Beheshti, I., Demirel, H., Matsuda, H., A.D.N.I., et al. (2017). Classification of Alzheimer's disease and prediction of mild cognitive impairment-to- Alzheimer's conversion from structural magnetic resource imaging using feature ranking and a genetic algorithm. *Computers in Biology and Medicine*, *83*, 109–119.

Benjamini, Y., Krieger, A.M., Yekutieli, D. (2006). Adaptive linear stepup procedures that control the false discovery rate. *Biometrika*, *93*(3), 491–507.

Bernal-Rusiel, J.L., Greve, D.N., Reuter, M., Fischl, B., Sabuncu, M.R., A.D.N.I., et al. (2013). Statistical analysis of longitudinal neuroimage data with linear mixed effects models. *Neuroimage*, *66*, 249–260.

Bernal-Rusiel, J.L., Reuter, M., Greve, D.N., Fischl, B., Sabuncu, M.R., A.D.N.I., et al. (2013). Spatiotemporal linear mixed effects modeling for the mass-univariate analysis of longitudinal neuroimage data. *Neuroim- Age*, *81*, 358–370.

Braak, H., & Braak, E. (1991). Neuropathological stageing of Alzheimerrelated changes. *Acta Neuropathologica*, *82*(4), 239–259.

Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehéricy, 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*(2), 766–781.

Da, X., Toledo, J.B., Zee, J., Wolk, D.A., Xie, S.X., Ou, Y., et al. (2014). Integration and relative value of biomarkers for prediction of MCI to AD progression: spatial patterns of brain atrophy, cognitive scores, APOE genotype and CSF biomarkers. *NeuroImage: Clinical*, *4*, 164–173.

Desikan, R.S., Cabral, H.J., Fischl, B., Guttmann, C.R., Blacker, D., Hyman, B.T., et al. (2009). Temporoparietal MR imaging measures of atrophy in subjects with mild cognitive impairment that predict subsequent diagnosis of Alzheimer disease. *American Journal of Neuroradiology*, *30*(3), 532–538.

Dickerson, B.C., Feczko, E., Augustinack, J.C., Pacheco, J., Morris, J.C., Fischl, B., et al. (2009). Differential effects of aging and Alzheimer's disease on medial temporal lobe cortical thickness and surface area. *Neurobiology of Aging*, *30*(3), 432–440.

Eskildsen, S.F., Coupé, P., Fonov, V.S., Pruessner, J.C., Collins, D.L. (2015). Structural imaging biomarkers of Alzheimer's disease: predicting disease progression. *Neurobiology of Aging*, *36*, S23–S31.

Eskildsen, S.F., Coupé, P., García-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.

Falahati, F., Westman, E., Simmons, A. (2014). Multivariate data analysis and machine learning in Alzheimer's disease with a focus on structural magnetic resonance imaging. *Journal of Alzheimer's Disease*, *41*(3), 685–708.

Ferreira, F.L., Cardoso, S., Silva, D., Guerreiro, M., Mendonça, A.d.e., Madeira, S.C. (2017). Improving prognostic prediction from mild cognitive impairment to Alzheimer's disease using genetic algorithms. In *International conference on practical applications of computational biol- ogy & bioinformatics* (pp. 180–188).

Gavidia-Bovadilla, G., Kanaan-Izquierdo, S., Mataró-Serrat, M., Perera-Lluna, A., A.D.N.I., et al. (2017). Early prediction of Alzheimer's disease using null longitudinal model-based classifiers. *PloS One*, *12*(1), e0168011.

Guerrero, R., Schmidt-Richberg, A., Ledig, C., Tong, T., Wolz, R., Rueckert, D., et al. (2016). Instantiated mixed effects modeling of Alzheimer's disease markers. *NeuroImage*, *142*, 113–125.

Iglesias, J.E., Van Leemput, K., Augustinack, J., Insausti, R., Fischl, B., Reuter, M., et al. (2016). Bayesian longitudinal segmentation of hippocampal substructures in brain MRI using subject-specific atlases. *Neu- roImage*, *141*, 542–555.

Jack, C.R.J.r. (2012). Alzheimer disease: new concepts on its neurobiology and the clinical role imaging will play. *Radiology*, *263*(2), 344–361.

Jack, J.r., C.R., Knopman, D.S., Jagust, W.J., Shaw, L.M., Aisen, P.S., Weiner, M.W., et al. (2010). Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *The Lancet Neurology*, *9*(1), 119–128.

Jiang, J., Sachdev, P., Lipnicki, D.M., Zhang, H., Liu, T., Zhu, W., et al. (2014). A longitudinal study of brain atrophy over two years in community-dwelling older individuals. *Neuroimage*, *86*, 203–211.

Korolev, I.O., Symonds, L.L., Bozoki, A.C., et al. (2016). Predicting progression from mild cognitive impairment to Alzheimer's dementia using clinical, MRI, and plasma biomarkers via probabilistic pattern classification. *PloS One*, *11*(2), e0138866.

LaMontagne, P.J., Keefe, S., Lauren, W., Xiong, C., Grant, E.A., Moulder, K.L., et al. (2018). Oasis-3: Longitudinal neuroimaging,

clinical, and cognitive dataset for normal aging and alzheimer's disease. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *14*(7), P1097.

Landin-Romero, R., Kumfor, F., Leyton, C.E., Irish, M., Hodges, J.R., Piguet, O. (2017). Disease-specific patterns of cortical and subcortical degeneration in a longitudinal study of Alzheimer's disease and behavioural-variant frontotemporal dementia. *Neuroimage*, *151*, 72–80.

Li, Y., Wang, Y., Wu, G., Shi, F., Zhou, L., Lin, W., et al. (2012). Discriminant analysis of longitudinal cortical thickness changes in Alzheimer's disease using dynamic and network features. *Neurobiology of Aging*, *33*(2), 427–e15.

Liu, Y., Julkunen, V., Paajanen, T., Westman, E., Wahlund, L.-O., Aitken, A., et al. (2012). Education increases reserve against Alzheimer's disease evidence from structural MRI analysis. *Neuroradiology*, *54*(9), 929–938.

Markesbery, W.R. (2010). Neuropathologic alterations in mild cognitive impairment: a review. *Journal of Alzheimer's Disease*, *19*(1), 221–228.

Minhas, S., Khanum, A., Riaz, F., Khan, S., Alvi, A. (2017). Predicting progression from mild cognitive impairment to Alzheimer's disease using autoregressive modelling of longitudinal and multimodal biomarkers. IEEE Journal of Biomedical and Health Informatics.

Moradi, E., Pepe, A., Gaser, C., Huttunen, H., Tohka, J., A.D.N.I., et al. (2015). Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *Neuroimage*, *104*, 398–412.

Park, H., Yang, J.-j., Seo, J., Lee, J.-m., et al. (2013). Dimensionality reduced cortical features and their use in predicting longitudinal changes in Alzheimer's disease. *Neuroscience Letters*, *550*, 17–22.

Peng, H., Long, F., Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*(8), 1226–1238.

Petersen, R.C., Roberts, R.O., Knopman, D.S., Boeve, B.F., Geda, Y.E., Ivnik, R.J., et al. (2009). Mild cognitive impairment: ten years later. *Archives of Neurology*, *66*(12), 1447–1455.

Pettigrew, C., Soldan, A., Zhu, Y., Wang, M.-C., Moghekar, A., Brown, T., et al. (2016). Cortical thickness in relation to clinical symptom onset in preclinical AD. *NeuroImage: Clinical*, *12*, 116–122.

Platero, C., Lin, L., Tobar, M.C. (2018). Longitudinal neuroimaging hippocampal markers for diagnosing Alzheimer's disease. Neuroinfor- matics. Retrieved from https://doi.org/10.1007/s12021-018-9380-2.

Rathore, S., Habes, M., Iftikhar, M.A., Shacklett, A., Davatzikos, C. (2017). A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages. NeuroImage.

Roberts, R., & Knopman, D.S. (2013). Classification and epidemiology of MCI. *Clinics in Geriatric Medicine*, *29*(4), 753–772.

Saykin, A.J., Shen, L., Foroud, T.M., Potkin, S.G., Swaminathan, S., Kim, S., et al. (2010). Alzheimer's Disease Neuroimaging Initiative biomarkers as quantitative phenotypes: genetics core aims, progress, and plans. *Alzheimer's & Dementia: the Journal of the Alzheimer's Association*, *6*(3), 265–273.

Sørensen, L., Igel, C., Pai, A., Balas, I., Anker, C., Lillholm, M., et al. (2017). Differential diagnosis of mild cognitive impairment and Alzheimer's disease using structural MRI cortical thickness, hippocampal shape, hippocampal texture, and volumetry. *NeuroImage: Clinical*, *13*, 470–482.

Thirion, B., Pinel, P., Mériaux, S., Roche, A., Dehaene, S., Poline, J.-B. (2007). Analysis of a large fMRI cohort: Statistical and methodological issues for group analyses. *NeuroImage*, *35*(1), 105–120.

Thompson, P.M., Hayashi, K.M., De Zubicaray, G., Janke, A.L., Rose, S.E., Semple, J., et al. (2003). Dynamics of gray matter loss in Alzheimer's disease. *Journal of Neuroscience*, *23*(3), 994–1005.

Weiner, M.W., & Veitch, D.P. (2015). Introduction to special issue: overview of alzheimer's disease neuroimaging initiative. *Alzheimer's & Dementia*, *11*(7), 730–733.

Westman, E., Aguilar, C., Muehlboeck, J.-S., Simmons, A. (2013). Regional magnetic resonance imaging measures for multivariate analysis in Alzheimer's disease and mild cognitive impairment. *Brain Topography*, *26*(1), 9–23.

Wolz, R., Heckemann, R.A., Aljabar, P., Hajnal, J.V., Hammers, A., Lötjönen, J., et al. (2010). Measurement of hippocampal atrophy using 4D graph-cut segmentation: application to ADNI. *NeuroImage*, *52*(1), 109–118.

Wyman, B.T., Harvey, D.J., Crawford, K., Bernstein, M.A., Carmichael, O., Cole, P.E., et al. (2013). Standardization of analysis sets for reporting results from ADNI MRI data. *Alzheimer's & Dementia: the Journal of the Alzheimer's Association*, *9*(3), 332–337.