Prediction of clinical and biomarker conformed Alzheimer's disease and mild cognitive impairment from multi-feature brain structural MRI using age-correction from a large independent lifespan sample

Binyin Li[a,b,d,*], Miao Zhang[c], Joost Riphagen[a], Kathryn Morrison Yochim[a], Biao Li[c], Jun Liu[b], David H. Salat[a,d,e], for the Alzheimer's Disease Neuroimaging Initiative[1]

[a] MGH/MIT/HMS Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School, Charlestown, MA, USA
[b] Department of Neurology, Ruijin Hospital Affiliated with Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China
[d] Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA
[c] Department of Nuclear Medicine, Ruijin Hospital Affiliated with Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China
[e] Neuroimaging Research for Veterans Center, VA Boston Healthcare System, Boston, MA, USA

ARTICLE INFO

ABSTRACT

Structural neuroimaging has been applied to the identification of individuals with Alzheimer's disease (AD) and mild cognitive impairment (MCI). However, these methods are greatly impacted by age limiting their utility for detection of preclinical pathology.

We built linear models for age based on multiple combined structural features using a large independent lifespan sample of 272 healthy adults across a wide age range from the Human Connectome Project Aging study. These models were then used to create a new support vector machine (SVM) training model in 136 CE and 268 control participants based on residues of fit from the expected age-effects relationship. Subsequent validation assessed the accuracy of the SVM model in new datasets. Finally, we applied the classifier to 276 individuals with MCI to evaluate prediction for early impairment and longitudinal cognitive change.

The optimal 10-fold cross-validation accuracy was 93.07%, compared to 91.83% without age detrending. In the validation dataset, the classifier for AD obtained an accuracy of 84.85% (56/66), sensitivity of 85.36% (35/41) and specificity of 84% (21/25). Classification accuracy was improved when using the lifespan sample as opposed to the classification sample. Importantly, we observed cross-sectional greater AD specific biomarkers, as well as faster cognitive decline in MCI who were classified as more 'AD-like' (MCI-AD), and these effects were pronounced in individuals who were late MCI. The top five contributive features were volumes of left hippocampus, right hippocampus, left amygdala, the thickness of left and right middle temporal & parahippocampus gyrus.

Linear detrending for age in SVM for combined structural features resulted in good performance for recognition of AD and AD-specific biomarkers, as well as prediction of MCI progression. Such procedures may be used in future work to enhance prediction in samples with atypical age distributions.

1. Introduction

Alzheimer's disease (AD) is the most common cause of dementia in older adults and is characterized by abnormal pathologic proteins (e.g. amyloid and tau) that are presumed to promote neural injury (Jack et al., 2018). Compared to pathological biomarkers of amyloid and tau

that can be visualized by positron emission tomography (PET) or quantified in cerebrospinal fluid (CSF), structural MRI is less costly, less invasive and available in most clinical settings. Thus, the utilization of structural MRI in the accurate and sensitive detection of AD or even AD pathology is an important goal.

Various structural features obtained from simple T1 weighted MRI

---

are impacted by typical aging as well as the degenerative processes of AD (Jefferson et al., 2015; Lindemer et al., 2015, 2017a; Coutu et al., 2016; Belathur Suresh et al., 2018) and such features have been used in statistical classification of individuals with a clinical diagnosis of AD (Chetelat and Baron, 2003; Teipel et al., 2013; Park and Moon, 2016; Belathur Suresh et al., 2018). The patterns of atrophy measured by computational models of cerebral cortical thickness can provide spatially distinct features to be utilized in classification models (Eskildsen et al., 2013; Raamana et al., 2015; de Vos et al., 2016). Similarly, subcortical volumetric atrophy can provide metrics that are sensitive to neurodegenerative processes in AD as well as the impact of typical aging, with the hippocampus, amygdala, and thalamus being particularly affected in AD (Gosche et al., 2002; Kwak et al., 2018). In addition to morphometric properties, additional 'microstructural' features can be obtained from standard structural MRI through quantification of signal properties such as the gray to white matter signal intensity ratio (GWR) at each point along the cortical surface. The GWR is regionally increased with age particularly throughout frontal regions demonstrating a reduction in the gray-white matter contrast (Salat et al., 2009). The GWR in temporal and limbic tissue is selectively associated with decreased hippocampal volume in AD (Salat et al., 2011). Additionally, white matter signal abnormalities (WMSAs; aka white matter hyperintensities) are presumed to be due to small vessel disease (Fazekas et al., 1988; Salat, 2014), more highly prevalent in individuals with AD (Lindemer et al., 2017a, 2017b; Nasrabady et al., 2018), and known to impact clinical trajectories as well as potentially modulate clinical status for a given level of primary AD pathology (Lindemer et al., 2018). In summary, these results demonstrate the range of structural features that can be obtained from standard T1 imaging and used in the classification of individuals with AD pathology as well as the trajectory of cognitive decline. However, given the overlapping nature of aging and AD-related effects on these features, careful modeling is required to assure that classifications are not influenced by the strong contributing factor. This is particularly true in the extremes of the age-range as the models may not have adequate data to model those segments appropriately.

Support vector machine (SVM) is currently the most widely used procedure in neuroimaging studies to classify AD (Cuingnet et al., 2011; Aguilar et al., 2013; Salvatore et al., 2016; Belathur Suresh et al., 2018). It is possible that such procedures using simple structural imaging can provide initial screening to determine whether follow up assessment with more direct, yet costly and invasive biomarker procedures are necessary. Given the impact of age on structural imaging measures, it is possible that age effects influenced prior results and need to be considered more carefully. For example, we previously found that SVM using regional structural imaging features that show statistical differences between AD and controls (entorhinal cortex, precuneus cortex, fusiform gyrus, banks of the superior temporal sulcus, inferior parietal cortex, supramarginal gyrus, superior frontal gyrus, inferior temporal gyrus, superior parietal cortex, superior temporal gyrus, middle temporal gyrus, rostral middle frontal gyrus, caudal middle frontal gyrus, and pars opercularis) resulted in older control individuals being more likely to be classified as AD (Belathur Suresh et al., 2018). Additionally, several prior studies have relied on clinical labels for the determination of classification accuracy. Important recent work describes the utility of a biological framework for AD (Jack et al., 2019) that includes information across major disease biomarker categories: β-amyloid (Aβ, A), pathological tau (T), and neurodegeneration (N). Thus, novel methods for classification must be assessed using this biomarker information as well as information about cognitive trajectories in addition to the clinical diagnosis for a full understanding of performance.

In this work, we implement linear detrending of structural MRI features for age in a large independent lifespan sample prior to SVM classification. We validated the classifier on a third independent dataset of AD and control participants. In addition, we applied the classifier to mild cognitive impairment (MCI) to evaluate its recognition for Aβ (by

^{18}F-florbetapir PET, labeled "A" biomarker), CSF phosphorylated tau (p-tau, labeled "T" biomarker) and neurodegeneration (by ^{18}F-fluorodeoxyglucose (FDG) PET, labeled "N" biomarker). Finally, we examined longitudinal cognitive trajectories in MCI to further validate the classification results.

2. Materials and methods

2.1. Dataset

2.1.1. Dataset one for linear detrending models (D1)

We used structural MRI data from 272 healthy adults in the Human Connectome Project Lifespan-Aging cohort (HCP-A) (146 women, age: 36 - > 100). T1 weighted multi-echo MPRAGE with prospective navigator motion correction MRI parameters included TE = 1.8/3.6/5.4/7.2 ms (multi-echo), TR = 2500 ms, field of view = 256 × 256 mm^2, number of slices = 208, voxel size = 0.8 × 0.8 × 0.8 mm^3, and flip angle = 8° (Harms et al., 2018; Bookheimer et al., 2019).

2.1.2. Dataset two for training and testing classifier (D2)

T1-weighted images from AD and control (CN) participants from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset (adni.loni.usc.edu) were used for training the classifier. A total of 404 participants (268 CN and 136 CE) were considered in this study similar to description in our prior work (Belathur Suresh et al., 2018). MRI was performed on 3.0 T scanners with the following parameters described by ADNI-2 protocols. Siemens (54% AD and 60% CN): 3D Magnetization Prepared-Rapid Gradient Echo (MPRAGE), TR = 2300 ms, TE = 2.98 ms, flip angle = 9°, voxel size = 1 × 1 × 1.2 mm^3. Philips (20% AD and 18% CN): 3D MPRAGE, TR = 6.8 ms, TE = 3.1 ms, flip angle = 9°, voxel size = 1 × 1 × 1.2 mm^3. GE (26% AD and 22% CN): 3D inversion-recovery spoiled gradient-recalled (IR-SPGR), TR = 2300 ms, TE = Min Full, flip angle = 11°, voxel size = 1 × 1 × 1.2 mm^3.

PET data were acquired using ^{18}F florbetapir (AV45) for imaging Aβ. The standardized uptake value ratios (SUVR) were calculated as the average of the uptake values of the frontal, angular/posterior cingulate, lateral parietal, and temporal cortices divided by the mean uptake values of the cerebellum, following a standardized pipeline (Landau et al., 2013). The Aβ positivity was derived using SUVR cut-off of 1.11 as described previously (Landau et al., 2014). Among the 136 CE and 268 CN included in our study, 119 of AD were Aβ + and 181 of CN were Aβ-. The 119 CE and 181 CN were defined as D2-plus for these confirmatory analyses (Table 1).

2.1.3. Dataset three for validation (D3)

We evaluated 66 participants using a whole-body PET/MR scanner (Biograph mMR; Siemens Healthcare, Erlangen, Germany) with an 8-channel head/neck coil in Shanghai Jiao Tong University affiliated Ruijin Hospital, China. The Aβ SUVR was calculated as D2 plus by ^{18}F florbetapir PET. The simultaneous MR models: MPRAGE, TR = 1900 ms, TE = 2.44 ms, angle = 9°, 192 slices (gap, 0.5 mm) covering the whole brain, FOV 256 × 256 mm, matrix 256 × 256, voxel size = 1.0 × 1.0 × 1.0 mm^3. All patients performed Mini-Mental State Examination (MMSE, Chinese Version) (Katzman et al., 1988), Beijing version of Montreal cognitive assessment (MoCA), the Chinese version of Addenbrooke's cognitive examination-revised (ACER) (Fang et al., 2014), and global clinical dementia rating. Forty-one participants were diagnosed as AD according to the National Institute on Aging-Alzheimer's Association (NIA-AA) workgroups (Jack et al., 2018). This part was approved by the Ethics Committee, Shanghai Jiao Tong University affiliated Ruijin Hospital, China. Notably, these participants differed in primary racial composition from the D1 and D2 datasets, and therefore provide additional cross-racial validation.

Table 1
Participants demographics from the three datasets.

	Dataset 1	Dataset 2 (ADNI)		Dataset 2-plus (ADNI)		Dataset 3	
	HCP	AD	CN	AD	CN	AD	CN
N	272	136	268	119	181	41	25
Sex (F/M)	146/126	57/79	148/120	55/64	87/94	19/22	12/13
Age (year)	62.7 ± 16.8	74.2 ± 8.2	72.9 ± 6.0	73.8 ± 8.2	71.9 ± 5.9	68.7 ± 9.0	68.5 ± 6.1
Education (year)	15.3 ± 4.5	15.7 ± 2.5	16. 6 ± 2.5	15.6 ± 2.5	16. 8 ± 2.4	12.3 ± 3.3	13.2 ± 3.1
MoCA	26.2 ± 2.6	17.2 ± 4.5	25.8 ± 2.4	17.2 ± 4.5	25.9 ± 2.5	16.2 ± 6.8	27.2 ± 1.8
MMSE	–	23.0 ± 2.1	29.1 ± 1.1	23.1 ± 2.1	29.1 ± 1.2	20.7 ± 6.2	27.9 ± 2.0
ACE-R	–	–	–	–	–	60.5 ± 21.9	84.6 ± 18.6

HCP, human connectome project; AD, Alzheimer's disease; CN, cognitively normal; F, female; M, male; MMSE, mini-mental state examination; MoCA, Montreal cognitive assessment; ACER, Addenbrooke's cognitive examination-revised.

2.1.4. MCI dataset

To determine the utility of the classification procedures with earlier disease, we examined 180 early MCI (EMCI) and 96 late MCI (LMCI) patients as defined by ADNI with complete baseline and follow-up neuropsychological assessments. Baseline imaging was performed as for D2. AV45 PET, FDG-PET and CSF p-tau assessments were used for biological validation of classification results (Jagust et al., 2015). We applied classifiers from D2 to the MCI dataset to evaluate predictive performance in this earlier disease sample.

We additionally examined whether MCI classified as AD had a different longitudinal trajectory in cognitive scores relative to MCI classified as control across available visits: baseline, month 6, month 12, month 24, month 36 and month 48. Patients were evaluated by MMSE, MoCA, Alzheimer's Disease Assessment Scale Cognitive Subscale (ADAS-13) (Mohs et al., 1997), Rey's auditory verbal learning test (RAVLT) learning and forgetting score, and trail making test-B (TMT-B) in each visit. Demographics for the entire study are presented in Table 1 and Table 2.

2.2. Structural features

The preprocessing pipeline of structural images was performed using FreeSurfer version 6.0 (https://surfer.nmr.mgh.harvard.edu/) (Dale et al., 1999; Fischl et al., 1999; Fischl and Dale, 2000). As described in our prior publications (Belathur Suresh et al., 2018), after spatial and intensity normalization and skull stripping, the resulting volume was segmented into grey matter, white matter, and CSF, and a deformable surface algorithm was used to identify the pial surface. Cortical thickness was determined by measuring the distance between the white matter and pial surfaces.

Subcortical segmentation volumes: Automatic subcortical segmentation of a brain volume was based upon the existence of an atlas containing probabilistic information on the location of structures (Fischl et al., 2002). Estimated total intracranial volume (eTIV) and volumes of 16 regions of interest were extracted, including cerebral white matter, cerebral cortex, thalamus, caudate, putamen, pallidum, hippocampus and amygdala in each hemisphere. The WMSAs were

labeled using a probabilistic procedure subsequently extended to label white matter lesions. Total WMSA (hypointensity) volume was then calculated for each hemisphere; these were averaged together to create a single WMSA volume for each participant.

Gray-white matter ratio (GWR): The gray matter values as a ratio to bordering white matter values provided a unit that was normalized for the local imaging environment. As described previously (Salat et al., 2009, 2011), tissue intensities were measured 30% through the thickness of the cortical ribbon, normal to the gray/white border for gray matter, and 1 mm subjacent to the gray/white border along the surface normal for white matter. The 30% sampling procedure was utilized to be conservatively close to the gray/white border and the white matter sampling voxel (which should minimize potential spurious effects that could arise in sampling from more remote locations) and was additionally able to adjust in regions of low cortical thickness (as opposed to using a constant value across the entire border which could be problematic for thinner cortical areas).

We extracted averaged cortical thickness and GWRs in 74 labels per hemisphere from the Destrieux atlas (Destrieux et al., 2010) respectively, and thus each participant had 296 features from the brain surface. The subcortical segmentation volumes (16 ROIs, eTIV, WMSA) were log-transformed because volume ratio data did not conform to a Gaussian distribution (Szabo et al., 2003). Finally, a total of 314 features for each participant were prepared for analysis.

2.3. Feature selection by linear detrending for age

Linear regression models were created between age and each structural feature from D1. We evaluated all linear regression models (n = 314), and preserved the coefficients (β) and offsets (x_0) only when the feature significantly correlated with age at $p < 0.05$ without correction.

$$feature = \beta \times age + x_0$$

Then ages from D2 and D3 were put into the corresponding linear models, and then we calculated the residual of fit as the difference between the actual feature and predicted one by coefficients (β) and
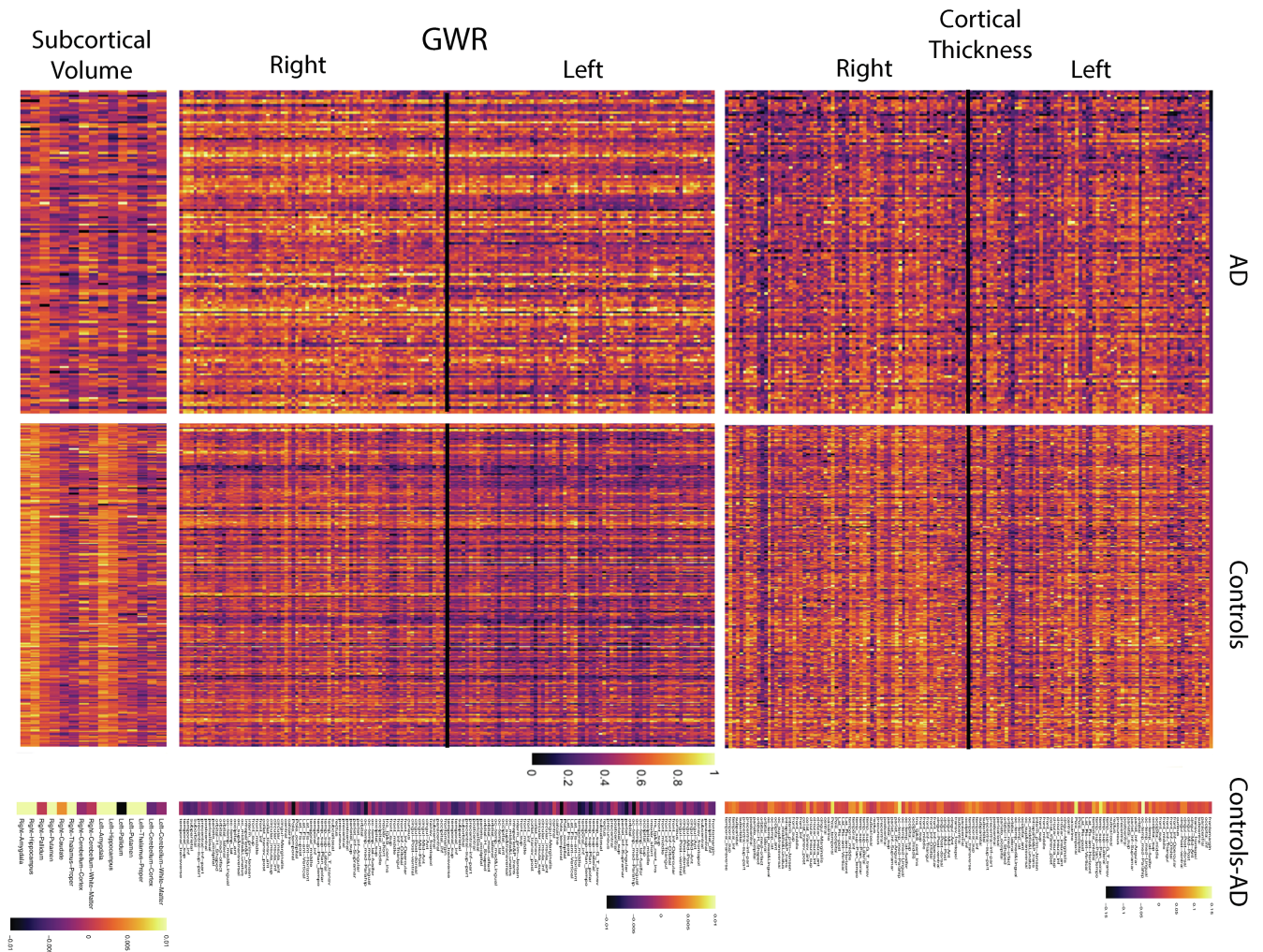
Table 2
Classification performance of models based on age detrending from different samples.

Model	Detrending model from D1	Detrending model from sub-D1	Detrending model from D2 control	Detrending model from D2-plus control
N	272	134	268	181
Age span	36 - > 100	55–85	56–90	56–89
Accuracy within SVM	392/404 (97.03%)	395/404 (97.77%)	384/404 (95.05%)	285/300 (95.00%)
Sensitivity within SVM	128/136 (94.12%)	129/136 (94.85%)	119/136 (87.50%)	107/119 (89.92%)
Specificity within SVM	264/268 (98.51%)	266/268 (99.25%)	265/268 (98.88%)	178/181 (98.34%)
Accuracy in D3	56/66 (84.84%)	53/66 (80.30%)	51/66 (77.27%)	54/66 (81.82%)
Sensitivity in D3	35/41 (85.36%)	33/41 (80.49%)	29/41 (70.73%)	32/41 (78.05%)
Specificity in D3	21/25 (84.00%)	20/25 (80.00%)	22/25 (88.00%)	22/25 (88.00%)

SVM, support vector machine.

Fig. 1. Different residuals from regression in AD and controls. Left: In each heatmap, columns represented participants, and rows represented regional difference between the actual measure and predicted one from age regression line (each row rescaled by max–min normalization across two groups, with the maximum = 1 and minimum = 0). The darker color showed relatively more negative residual, while lighter color suggested more positive residual when rows were compared between AD and controls from D2. Right: Averaged difference between value of each row in two groups from the left heatmap (normal controls - AD). The regions that the rows represented were marked on the right. AD, Alzheimer's disease; GWR, ratio of gray to white matter signal intensity.

offsets (x_0) in each model. The following m is the number of significant models evaluated above.

$$\varepsilon_m = \text{feature}_m - (\beta_m \times \text{age} + x_{m0})$$

After detrending, each participant had a vector of m residuals which represented the features without linear age effect. We had a $136 \times m$ matrix for the AD group and $268 \times m$ matrix for controls in D2, as well as $66 \times m$ matrix for all participants in D3. For MCI, we had $180 \times m$ matrix for EMCI and $96 \times m$ matrix for LMCI from their baseline MRI.

2.4. Support vector machine classifier

Support vector machine (SVM) is a commonly utilized supervised, multivariate classification method. The SVM classifier finds a hyperplane maximizing the margin between groups. The problem of AD detection in D2 using SVM was formulated as a binary classification problem. In this study, we used the SVM implementation publicly available in LibSVM (csie.ntu.edu.tw/cjlin/libsvm). The cost parameter C and kernel parameter γ of the nonlinear Gaussian function in the SVM classifier were optimized using cross-validation via the grid-search approach (Chang and Lin, 2011). The grid search was performed over the ranges C = 2^{-5}, 2^{-4}, …2^{15}, γ = 2^{-15}, 2^{-14}, …, 2^{5}. The optimized set of parameters was then used to train the SVM classifier by the input of

features after linear detrending from D2. We used the 10-fold cross-validation to obtain an estimate of the classifier performance. During each fold the classifier was developed using data from 90% of the participants and tested using data from the remaining 10% of the participants.

In this study, we adopted the F-score method in the libsvm feature selection tool to further evaluate feature contribution for classification. The F-score method has been generally used in pattern recognition systems to select the optimal feature subset (Polat and Güneş, 2009; Chen and Li, 2010). A larger F-score value indicates that the feature has more discriminative power. For validation, we used the SVM classifier with optimized parameters to predict each participant in D3, EMCI and LMCI by using their residual matrices mentioned above. The overall accuracy, specificity and sensitivity were reported as the generalization of the classifier. We also compared the results by including only "cortical thickness", "GWR", "subcortical volumes" or "WMSA" into the SVM classifier to determine performance for individual structural classes.

2.5. Statistics

Statistical analyses were performed with R (Version 3.6.2). Group differences were assessed using independent t-tests for continuous and
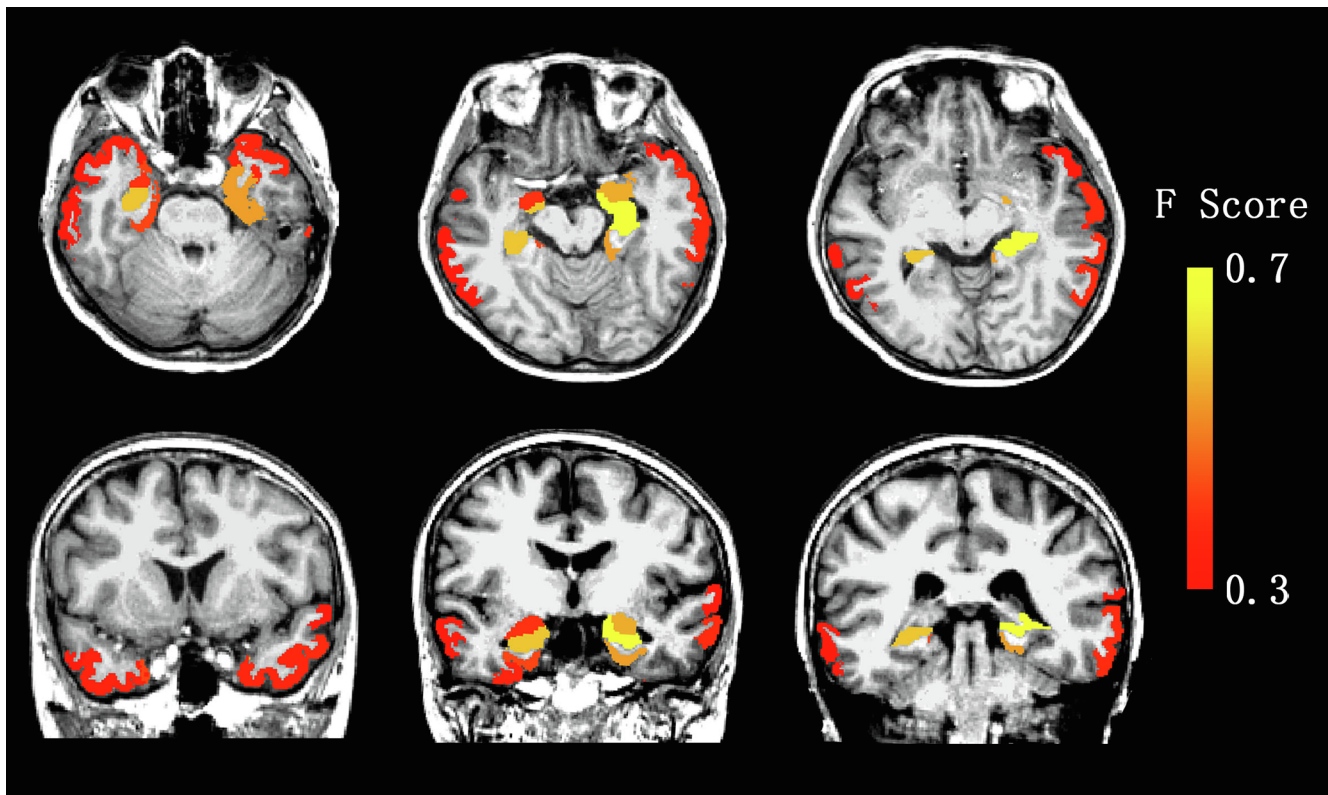
Fig. 2. Mapping of most contributive features for classification. Maps of subcortical volumes and cortical thickness with F-score > 0.3, in the SVM classifier from the controls and AD groups in D2.

Chi-squared tests for nominal data. All t-tests were assessed by the "RVAideMemoire" package in R, with 10,000 permutations for multiple comparison correction. We used a generalized linear model to control age, sex and education effect in the MCI baseline comparison. To compare the longitudinally cognitive changes in MCI, we performed a mixed model procedure for repeated measures, and evaluated the interaction between classified group and follow-up timelines. All tests were 2 tailed, and values of corrected p < 0.05 were considered statistically significant unless specified.

3. Results

3.1. Linear detrending for age in D1

Among the 148 surface parcellations, cortical thickness in 10 regions was not significantly correlated with age and excluded from detrending, including middle-anterior part of the cingulate gyrus and sulcus, long insular gyrus and central sulcus of the insula, inferior temporal gyrus, anterior transverse collateral sulcus in the left hemisphere, fronto-marginal gyrus and sulcus, cuneus, orbit gyrus, and pericallosal sulcus in the right hemisphere, and bilateral short insular gyri. In the 16 subcortical segmentations, only left caudate was excluded from the first-level selection. The GWR in 148 parcellations were all found correlated with age and included, in addition to WMSA and eTIV. The p values for each regression here were not corrected for multiple comparisons as the primary goal of this analysis was to determine age trends for each structure. In all, 303 age-related features were found, and their coefficients with age and corresponding offsets were used for calculating residuals of the fit in AD and controls.

3.2. Patterns of residual in D2

The residual matrix of AD was 136 × 303, and that of controls was

268 × 303 (138 features in thickness, 148 in GWRs, 15 in subcortical volumes, WMSA and eTIV). In the heatmap, each column represented one participant, and each row represented regional difference between the actual measure and predicted one from age regression best-fit line (Min-max normalization across two groups, Fig. 1). Among 138 surface regions, as expected, cortical thickness of AD deviated (lower residuals) more negatively than that of controls in 126 regions (mean error difference: − 0.478 to − 0.029, p < 0.05), while the residuals were similar in the remaining 21 regions. AD also deviated more negatively in the age-subcortical volume best-fit line in 8 regions, except left pallidum. AD had more positive residuals from the age-GWR fit line in 129 surface regions (mean error difference: 0.003 to 0.005, p < 0.05) and from age-WMSA line as well (Tables S1–S3).

3.3. Performance of SVM classifier

To illustrate the performance of the SVM classifier, residuals in 138 parcellations of mean cortical thickness, 148 GWRs, 15 subcortical volumes, WMSA and eTIV were extracted from the 404 participants in D2 and used as features. A total of 303 features were used for classification. The sensitivity, specificity, and accuracy were calculated to measure the performance of the SVM classifier. Sensitivity is defined as the proportion of true positives that are correctly identified by the test and specificity is defined as the proportion of true negatives that are correctly identified by the test. Accuracy is calculated as the proportion of true results (both true positives and true negatives) by the test. In D2, the optimal cross-validation accuracy of 93.07%, total accuracy of 97.03%, sensitivity of 94.12% (128/136), and specificity of 98.51% (264/268) were obtained using the SVM classifier. Without linear detrending for the age before SVM, the optimal cross-validation accuracy was 91.83%, total accuracy 96.29%, sensitivity 91.91% (125/136), and specificity 98.51% (264/268).

The top 5 features whose F-score values were > 0.4 were volumes of

left hippocampus (F-score, 0.696), right hippocampus (0.605), left amygdala (0.572), thickness of left (0.550) and right (0.414) middle temporal & parahippocampus gyrus. The volume of right amygdala (0.382), left middle temporal gyrus thickness (0.366), left and right temporal pole thickness (0.356), and volume of left superior temporal gyrus (0.311) ranked from 6 to 10 (Fig. 2). The greatest impacts from GWR were in left (0.275) and right (0.228) medial temporal & para-hippocampus gyrus.

The optimal cross-validation accuracy decreased to 90.84% when only cortical thickness was included in the D2 SVM model, to 84.65% when only GWR included, to 89.10% with only subcortical volumes, and to 83.17% with only WMSA. Similarly, the total accuracy also decreased by 0.74% (3/404), 3.47% (14/404), 6.93% (28/404), 13.86% (56/404) with only one of the above measures in the D2 SVM model.

In D3, we had 41 ADs with positive β amyloid deposition and poor cognitive performance and 25 controls with negative β amyloid and normal cognition. When we applied the classifier to D3 for validation, the classifier reported an accuracy of 84.85% (56/66), sensitivity of 85.36% (35/41) and specificity of 84.00% (21/25). Similar to D2, overall accuracy also decreased when only one of these structural measures included in the classifier, by 3.03%, 15.1%, 9.10%, 21.21% respectively.

We repeated the SVM based on D2-plus, which was only composed of AD with Aβ + and CN with Aβ-. Within the SVM, the optimal accuracy was 99.00%, sensitivity 98.32% (117/119), and specificity 99.45% (180/181) demonstrating the superior performance when taking biomarkers into account for classification. In D3, the new classifier reported an accuracy of 84.85% (56/66), sensitivity of 95.12% (39/41) and specificity of 68.00% (17/25). As ADNI participants were scanned by three scanner vendors, we did vendor-wise normalization to minimize potential cohort effects. The normalization did not greatly improve classifier performance and caused lower accuracy when we applied to D3 (accuracy: 75–80%, Table S4). Thus, we used raw data derived classifiers for the following validation.

We also examined and compared age detrending models based on different samples (D1, D1 with age 55–85, D2 controls, and D2-plus controls). The detrending model based on the full set of D1 which was widely distributed on the age range led to better classification performance and generalizability in D3 (Table 2).

3.4. Prediction of MCI by the classifier

Based on D2 derived classifier, in the 180 EMCI, 29 patients were classified as AD (16.1%) (EMCI-AD), and in the 96 LMCI, 41 were classified as AD (42.7%) (LMCI-AD). The classified groups were similar in age, sex and education years. The EMCI-AD and LMCI-AD both had worse performance in MoCA and ADAS-13 at baseline compared with EMCI and LMCI classified as control (EMCI-CN, LMCI-CN) respectively ($p < 0.001$, after adjusting for age). LMCI-AD additionally performed worse in baseline MMSE, learning of RAVLT and TMT-B ($p = 0.006$, 0.009, and 0.015 after adjusting for age, Fig. 3). For ATN assessments, EMCI-AD and LMCI-AD both had significantly lower FDG SUVR (EMCI: β = −0.52, p = 0.008; LMCI: β = −1.06, p < 0.001, Table 3) after controlling for age, sex and education. Moreover, LMCI-AD had higher amyloid SUVR (β = 0.84, p < 0.001) and CSF p-tau level (β = 0.69, p = 0.002).

Longitudinally, EMCI-AD had a greater decline in ADAS-13 and learning of RAVLT by month 36 compared to EMCI-CN, and LMCI-AD had a more significant decrease in ADAS-13 and MoCA by month 24 (Table 2). To perform a more conservative analysis further minimizing age effects in the classification results, we examined only age and education matched EMCI-CN/AD and LMCI-CN/AD cases. Similar effects of classification were found on cognition measured by MoCA and ADAS-13 by month 48 in EMCI and MoCA, MMSE and ADAS-13 by month 24 in LMCI (Table 4).

Based on the D2-plus derived classifier, in the 180 EMCI, 32 patients were classified as EMCI-AD (17.8%), compared to 29 (16.1%) were classified as as EMCI-AD from D2 derived classifier. In the 96 LMCI, 43 (44.8%) were classified as LMCI-AD compared to 41 (42.7%) classified as LMCI-AD using the D2 derived classifier. Similar to D2 derived classifier, EMCI-AD and LMCI-AD had lower FDG SUVR (EMCI: β = −0.59, p < 0.002; LMCI: β = −1.25, p < 0.001). Moreover, LMCI-AD had higher amyloid SUVR (β = 0.56, p = 0.007) and borderline greater CSF p-tau level (β = 0.69, p = 0.08).

4. Discussion

Given the correspondence between patterns of cerebral atrophy measured by MRI and the pathological processes of AD, it is theoretically possible to use these cost-effective and accessible measurements in the early detection of individuals with AD neuropathology. Prior studies have identified regional atrophy patterns in patients with AD (Falahati et al., 2014; de Vos et al., 2016), and subsequently used as features in machine learning models for patient classification (Belathur Suresh et al., 2018). Although applied effectively in prior work, advances towards greater generalization of models are necessary. Here we explored the impact of different types of age correction on models validated using AD biomarkers and longitudinal cognitive trajectories. The classifiers showed improved performance based on using a large independent lifespan dataset for age correction. Although performance decreased in an additional independent validation dataset with a substantially different racial composition, the classifier retained high accuracy demonstrating the robustness of the procedures implemented.

4.1. Age correction approaches

Age is one of the most consequential factors limiting generalizability in classification of AD based on brain imaging features. In prior studies, younger AD patients were more likely to be misclassified as controls, as smaller age-related changes may have masked disease effects, while older healthy brains were more likely to be misclassified as patients due to age-related atrophy (Dukart et al., 2011; Falahati et al., 2016). Thus, it is possible that enhanced forms of age correction could benefit classification procedures when test cases are outside of the typical age range. Age may also partially mask other disease-related factors such as downstream effects of ApoE genotype, global cognitive impairment and sex (Falahati et al., 2016). Sex may be less impactful for measures used here. Although sex differences in cortical thickness are reported during adolescence, these differences are less apparent in later ages (Sowell et al., 2007; Gennatas et al., 2017). Thus, we primarily focused on correction for age effects in this work to enhance classification performance.

The linear model was widely used for estimating the relationship between age and brain features. Different to age prediction studies, chronological age was used as a predictor and assumed of no errors. It is unlikely that non-Gaussian distribution of predictor would cause much bias in the linear model framework (Smith et al., 2019).

In our age-feature models, we used linear fits from an independent lifespan sample. The D1 cohort had the widest age-span (from 36 to > 100) and was informative for mapping the full age trajectory. Our results supported that the detrending coefficients from D1 worked best for the classifier compared to use of control participants from the D2 sample. These results demonstrated the benefit of more accurate age modeling achieved from a large lifespan sample.

4.2. Feature selection for AD recognition

We obtained mean cortical thickness and GWR from anatomical parcellations instead of voxel or vertex, to minimize and smooth random variations from multi-site measurements. The features that greatly contributive to our classifier were similar to 'AD signature'
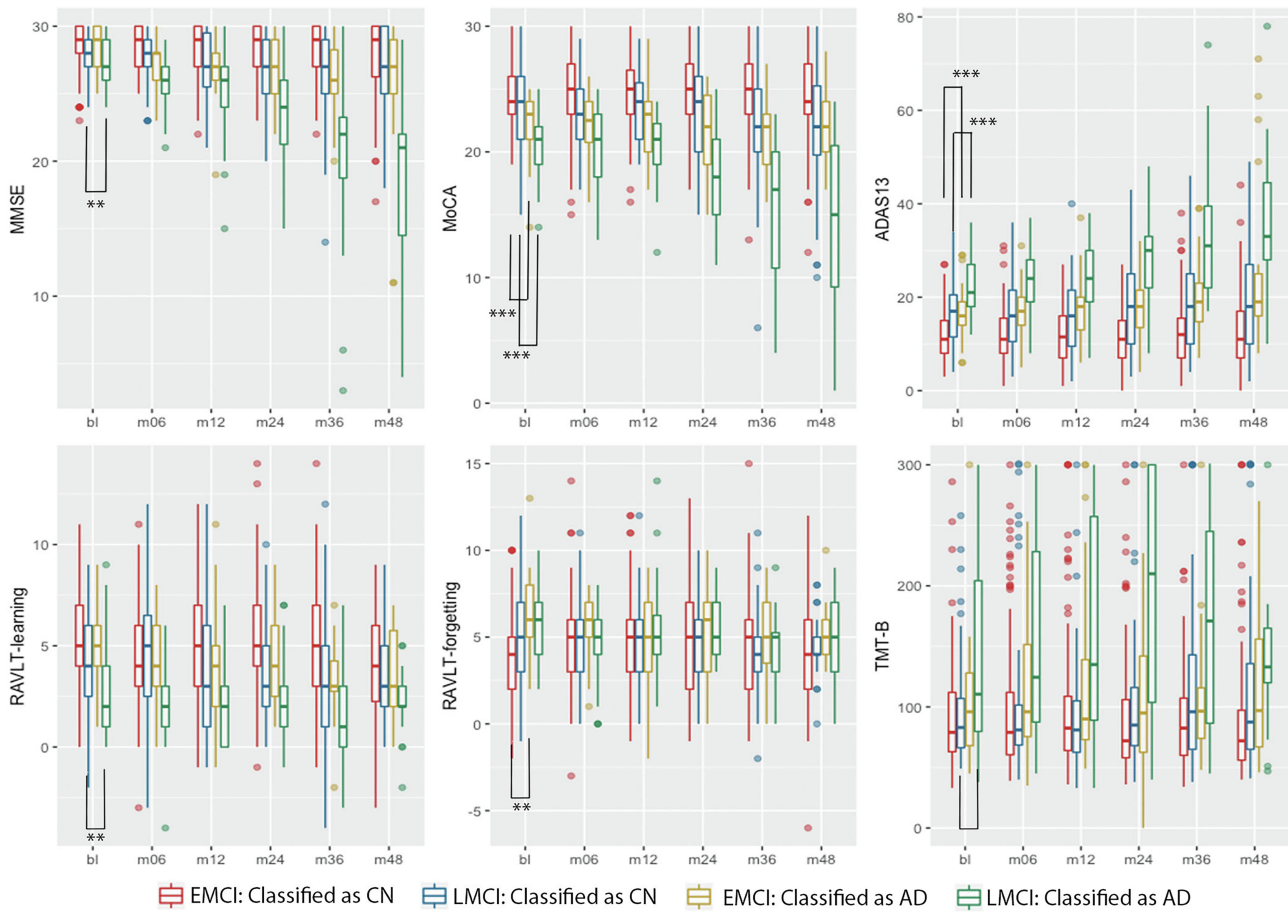
Fig. 3. Longitudinal changes in classified MCI. Boxplot of cognitive performance at each visit for MCI. MMSE, mini-mental state examination; MoCA, Montreal cognitive assessment; ADAS: Alzheimer's Disease Assessment Scale Cognitive Subscale; RAVLT, Rey's auditory verbal learning test; TMT-B, trail making test-B; EMCI, early mild cognitive impairment; LMCI, late mild cognitive impairment; CN, control; bl, baseline; m, month. *** p < 0.001, ** p < 0.01.

(Schwarz et al., 2016). Moreover, it is suggested that MRI-based morphometric estimates (cortical thickness in 'AD signature' regions, hippocampal volume and global atrophy) parallel CSF neurofilament light chain protein in differentiating individuals across the AD continuum on neurodegeneration status (Allison et al., 2019). Voxel-wise information could be preferable to summary features directly from the image as opposed to summary measures and given that voxel maps have been demonstrated to generalize well (Kloppel et al., 2008). However, we were able to achieve high accuracy across samples with generally accessible automated features, as well as decrease the likelihood of overfitting from large feature/sample ratio (Baumes et al., 2006). Ongoing work will examine the costs and benefits of the voxel compared to summary features.

The present study explored the way of merging brain structural features for detecting AD and predicting MCI progression. Among the features we included, the most prominent classification power was provided through volumes of hippocampus and amygdala. The results were consistent with observations from clinical practice and several prior research studies demonstrating early and profound atrophy in these regions with MCI and AD (Fischl et al., 2002; Desikan et al., 2009; Wachinger et al., 2016; Tentolouris-Piperas et al., 2017; Sengoku, 2019). In typical AD, medial temporal atrophy affects the amygdala and the hippocampus is usually accompanied by temporal horn enlargement and higher hazard ratio for disease progression (Desikan et al., 2010). Stronger atrophy in the amygdala and hippocampus predicted conversion to AD, and the linear discriminant analysis on the principal component values of hippocampus, amygdala, and ventricular volume provided classification 86%–88% for cognitive impairment (Tang et al.,

2014).

We had cross-validation accuracy of 89.1% with only subcortical features included, however, the accuracy increased with the addition of cortical thickness. Cortical thickness measures provided essential information for enhancing performance, especially in bilateral middle temporal and parahippocampal gyrus, as well as temporal pole. The pattern of cortical atrophy was partially consistent with the relationship between cortical thickness and cognition in AD dementia (Ossenkoppele et al., 2019). Given the relative accessibility of the features used in this work, the inclusion of the full feature set did not limit general utilization of the procedures.

GWR was reported as age- and AD-related (Salat et al., 2009, 2011; Westlye et al., 2009; Grydeland et al., 2013), while its value for classification was unclear. We previously examined the utility of GWR in identification of MCI progression (Jefferson et al., 2015). In the present study, the GWR measures improved accuracy compared to the structural features alone and the specificity increased in the classifier with only GWRs, while overall accuracy and sensitivity decreased. Compared with cortical thickness and subcortical volumes, the GWRs also had much lower F-scores. In all, we suggested GWRs less sensitive than other structural measures, but still helpful due to its specificity for AD classification.

White matter lesions measured as WMSA on MRI are commonly seen in normal aging and AD (de Leeuw et al., 2001), and histopathological studies have indicated a mix of heterogeneous findings that correlate with WMSA including demyelination and gliosis (Gouw et al., 2011). Individuals with AD have a higher total and greater regional volumes of WMSA regardless of age (Lindemer et al., 2015, 2017a,

Table 3
The classification of MCI and longitudinally cognitive changes.

	EMCI			LMCI		
	AD	CN	P-value	AD	CN	P-value
N	29	151	–	41	55	–
Age	72.9 ± 6.0	70.2 ± 6.9	0.031	73.3 ± 6.4	69.0 ± 7.3	0.002
Sex (F/M)	12/17	66/85	0.978	23/18	26/29	0.516
Amyloid PET[a]	1.2 ± 0.2	1.2 ± 0.1	0.861[#]	1.2 ± 0.2	1.4 ± 0.2	< 0.001[#]
FDG PET[b]	1.2 ± 0.1	1.3 ± 0.1	0.008[#]	1.1 ± 0.1	1.3 ± 0.1	< 0.001[#]
p-tau (ng/l)	24.4 ± 13.7	25.6 ± 14.9	0.518[#]	39.8 ± 15.0	27.7 ± 14.4	0.002[#]
Education	16.5 ± 3.0	16.5 ± 2.6	0.407	16.9 ± 2.6	16.2 ± 2.6	0.167

	EMCI[#]			LMCI[#]		
	Interaction* (p-value)	Effect estimate (Beta)	95% CI of Beta	Interaction* (p-value)	Effect estimate (Beta)	95% CI of Beta
ADAS-13	m36 (0.020)	3.60	1.25–5.95	m24 (< 0.001)	5.07	2.10–8.03
	m48 (< 0.001)	8.84	6.38–11.3	m36 (< 0.001)	9.32	6.10–12.5
				m48 (< 0.001)	11.2	7.64–14.8
MMSE	m06 (0.003)	−1.32	−2.20 to −0.45	m24 (0.002)	−2.00	−3.30 to −0.71
	m12 (0.006)	−1.21	−2.08 to −0.33	m36 (< 0.001)	−4.71	−6.12 to −3.31
	m24 (0.004)	−1.31	−2.20 to −0.41	m48 (< 0.001)	−6.55	−8.11 to −4.98
	m36 (< 0.001)	−2.10	−3.02 to −1.17			
	m48 (< 0.001)	−2.33	−3.29 to −1.36			
MOCA	–	–	–	m24 (0.001)	−2.25	−3.63 to −0.88
				m36 (< 0.001)	−3.62	−5.10 to −2.13
				m48 (< 0.001)	−4.49	−6.16 to −2.80
RAVLT-learning	m36 (0.047)	−1.14	−2.26 to −0.01	–	–	–
RAVLT-forgetting	m12 (0.002)	−1.92	−3.16 to −0.67	–	–	–
	m24 (0.013)	−1.61	−2.88 to −0.33			
	m36 (0.013)	−1.69	−3.02 to −0.35			
TMT-B	–	–	–	m12 (0.024)	27.3	3.62–51.1
				m24 (< 0.001)	57.6	33.0–82.2
				m36 (< 0.001)	48.8	21.5–76.0
				m48 (0.030)	35.8	3.51–68.0

* Interaction between follow-up and classification, with baseline as reference, only shown with p < 0.05.
Data are given in mean values (standard deviation, SD), if not otherwise specified. m, month of follow-up; M, male; F, female. AD, Alzheimer' disease; CN, cognitively normal; EMCI, early MCI; LMCI, late MCI; N, number of subjects; p-tau, phosphorylated tau in cerebrospinal fluid; MMSE, mini-mental state examination; MoCA, Montreal cognitive assessment; ADAS: Alzheimer's Disease Assessment Scale Cognitive Subscale; RAVLT, Rey's auditory verbal learning test; TMT-B, trail making test-B.
[a] Global neocortical uptake relative a composite reference region in florbetapir PET.
[b] Global neocortical uptake relative a composite reference region in fluorodeoxyglucose PET.
[#] Adjusted for age, sex and years of education.

2017b) and white matter lesions are associated with more rapid cognitive decline across time in MCI and AD (Coutu et al., 2017). In the present study, we found that the residual of total WMSA from age-regression line only had a minor impact on the SVM classifier (F-score 0.023). This may be due to the fact that WMSA volume is highly correlated with other prominent features of AD such as hippocampal

Table 4
Cognitive changes in age and education matched group for EMCI-AD and LMCI-AD.

	EMCI			LMCI		
	AD	CN	P-value	AD	CN	P-value
N	29	29	–	41	41	–
Age	72.9 ± 6.0	73.6 ± 7.2	0.702	73.3 ± 6.4	71.6 ± 6.0	0.234
Sex (Female/Male)	12/17	12/17	0.978	23/18	18/23	0.377
Education	16.5 ± 3.0	15.9 ± 2.6	0.406	16.9 ± 2.1	16.7 ± 2.4	0.735
MMSE-Baseline	28.3 ± 1.6	28.3 ± 1.6	0.933	27.0 ± 1.7	28.0 ± 1.7	0.007
Follow up*	Ns		–	m24		0.003
				m36		< 0.001
				m48		< 0.001
MOCA-Baseline	21.9 ± 2.4	24.5 ± 2.7	< 0.001	20.6 ± 2.5	23.6 ± 3.1	< 0.001
Follow up*	m48		0.054	m24		0.008
				m36		< 0.001
				m48		< 0.001
ADAS-Baseline	17.0 ± 5.9	12.0 ± 5.0	< 0.001	22.3 ± 5.8	16.4 ± 6.3	< 0.001
Follow up*	m48		0.007*	m24		< 0.001
				m36		< 0.001
				m48		< 0.001

*Significant interaction between follow-up and classification, with baseline as reference; m, month of follow-up. AD, Alzheimer's disease; CN, cognitively normal; EMCI, early MCI; LMCI, late MCI; N, number of subjects.

volume (Coutu et al., 2016).

4.3. Classifier optimization and validation

In the SVM, we set the residuals from the best fit lines as potential features as input. Thereby, age-related effects on MRI data were minimized and we focused on how far a person deviated from age-predicted measures (the residuals of the age regression for each feature). Although relatively successful for age correction, reviews of prior work demonstrated SVM model accuracy ranging from 85 to 91% (Dukart et al., 2011; Falahati et al., 2016; Belathur Suresh et al., 2018). Cortical thickness in AD signature regions, hippocampus volume and global atrophy discriminated between AD and controls with sensitivity from 80 to 90%, specificity from 85 to 90% (Allison et al., 2019). Using age-detrended features, we had an increase in model accuracy of 2–8% (mean cross-validation accuracy: 93.1%, total accuracy: 97.0%, sensitivity: 94.1%, specificity: 98.5%). Critically, using a novel dataset with substantial differences from the training dataset (differing in clinical sample, racial composition, and scanner type and imaging protocols), a biomarker confirmed accuracy of 84.8% was achieved demonstrating generalization at least across these conditions.

We created a biomarker-based classifier using only patients positive for Aβ and controls negative for Aβ which in theory should provide more pure classification based on pathology. This biomarker classifier had the same accuracy, higher sensitivity (95.1%) and lower specificity (68.0%) than the classifier based on clinical diagnosis. The clinically diagnosed D2 from ADNI-2 promoted a more balanced sensitivity and specificity while the amyloid confirmed D2-plus may be more useful as a sensitive tool in the clinical setting.

The critical validation in the present study was provided through examination of the cross-sectional ATN biomarkers and longitudinal cognitive trajectories in the sample of MCI. The EMCI-AD had lower FDG-PET metabolism. In the ADNI-2, LMCI had more severe episodic memory impairment than EMCI (Aisen et al., 2015). Among these LMCI, our classifiers (both from D2 and D2-plus) could find individuals with higher level of amyloid accumulation, tau pathology and neurodegeneration. In line with the predictive role of AD pathology in cognitive decline (Varatharajah et al., 2019), we could find the MCI-AD that would decline faster. The EMCI-AD and LMCI-AD were older than patients classified as controls. This may still be expected given the increase in prevalence of AD with increasing age and given potential age interactions with this condition (Yao et al., 2012). We conservatively performed an additional matching of the MCI groups and still observed significantly greater cognitive decline in both EMCI-AD and LMCI-AD compared to matched MCI classified as more control-like. Moreover, the LMCI-AD showed longitudinal differences earlier than EMCI-AD supporting their later stage of impairment.

4.4. Limitations

Several limitations of the current work are being explored in ongoing research. First, we corrected age effect based on the cross-sectional dataset and the relationship between structural features and age could be ambiguous due to subject variability. Although linear regression is widely used in age-brain analysis, non-linear detrending for age could also be considered with longitudinal datasets. Second, although cross-sectional data perform well as described in the manuscript, they are limited in not providing within individual change measures, and longitudinal measures could be more sensitively trained to expected trajectories across time which would have important prognostic value. Third, we could not fully exclude the "healthy" older adults in D1 who may also have some underlying AD pathology. However, the similar accuracy from D2 and D2-plus provides some confidence that to some extent structure from cognitively normal controls could be regarded as that from Aβ- controls. Finally, improvement for the proposed method could involve the inclusion of clinical variables that are also associated with structural measures in the linear detrending model prior to SVM.

5. Conclusions

The major conclusions from current work are as follows: (a) Linear detrending for age and followed SVM for combined structural measures provided good performance in AD recognition. (b) Compared to other structural measures, the volumes of hippocampus and amygdala, as well as the cortical thickness in medial and superior temporal gyrus, had more important influence on classification accuracy after removal of age effect. (c) The classifier could help recognition of cross-sectional ATN biomarkers, and prediction of longitudinal cognitive decline in MCI. Future work will examine the degree to which both cross-sectional and longitudinal pathology can be determined from more specific structural features.

CRediT authorship contribution statement

Binyin Li: Conceptualization, Methodology, Investigation, Formal analysis, Writing - original draft, Writing - review & editing. **Miao Zhang:** Conceptualization, Methodology, Investigation. **Joost Riphagen:** Investigation, Writing - review & editing. **Kathryn Morrison Yochim:** . **Biao Li:** Investigation. **Jun Liu:** Investigation. **David H. Salat:** Investigation, Formal analysis, Writing - review & editing. **:** .

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.nicl.2020.102387.

References

Aguilar, C., Westman, E., Muehlboeck, J.S., Mecocci, P., Vellas, B., Tsolaki, M., et al., 2013. Different multivariate techniques for automated classification of MRI data in Alzheimer's disease and mild cognitive impairment. Psychiatry Res. 212 (2), 89–98.

Aisen, P.S., Petersen, R.C., Donohue, M., Weiner, M.W., 2015. Alzheimer's disease neuroimaging I. Alzheimer's disease neuroimaging initiative 2 clinical core: progress and plans. Alzheimers Dement 11 (7), 734–739.

Allison, S.L., Koscik, R.L., Cary, R.P., Jonaitis, E.M., Rowley, H.A., Chin, N.A., et al., 2019. Comparison of different MRI-based morphometric estimates for defining neurodegeneration across the Alzheimer's disease continuum. Neuroimage Clin. 23, 101895.

Baumes, L.A., Serra, J.M., Serna, P., Corma, A., 2006. Support vector machines for predictive modeling in heterogeneous catalysis: a comprehensive introduction and overfitting investigation based on two real applications. J. Comb. Chem. 8 (4), 583–596.

Belathur Suresh, M., Fischl, B., Salat, D.H., 2018. Alzheimer's Disease Neuroimaging I. Factors influencing accuracy of cortical thickness in the diagnosis of Alzheimer's disease. Hum. Brain Mapp. 39 (4), 1500–1515.

Bookheimer, S.Y., Salat, D.H., Terpstra, M., Ances, B.M., Barch, D.M., Buckner, R.L., et al., 2019. The Lifespan Human Connectome Project in Aging: An overview. Neuroimage 185, 335–348.

Chang, C.C., Lin, C.-J., 2011. LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST).

Chen, F.-L., Li, F.-C., 2010. Combination of feature selection approaches with SVM in credit scoring. Expert Syst. Appl. 37 (7), 4902–4909.

Chetelat, G., Baron, J.C., 2003. Early diagnosis of Alzheimer's disease: contribution of structural neuroimaging. Neuroimage 18 (2), 525–541.

Coutu, J.P., Goldblatt, A., Rosas, H.D., Salat, D.H., 2016. Alzheimer's Disease Neuroimaging I. White Matter Changes are Associated with Ventricular Expansion in Aging, Mild Cognitive Impairment, and Alzheimer's Disease. J. Alzheimers Dis. 49 (2), 329–342.

Coutu, J.P., Lindemer, E.R., Konukoglu, E., Salat, D.H., 2017. Alzheimer's Disease Neuroimaging I. Two distinct classes of degenerative change are independently linked to clinical progression in mild cognitive impairment. Neurobiol. Aging 54, 1–9.

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 (2), 766–781.

Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage 9 (2), 179–194.

de Leeuw, F.E., de Groot, J.C., Achten, E., Oudkerk, M., Ramos, L.M., Heijboer, R., et al., 2001. Prevalence of cerebral white matter lesions in elderly people: a population based magnetic resonance imaging study. The Rotterdam Scan Study. J. Neurol. Neurosurg. Psychiatry 70 (1), 9–14.

de Vos, F., Schouten, T.M., Hafkemeijer, A., Dopper, E.G., van Swieten, J.C., de Rooij, M., et al., 2016. Combining multiple anatomical MRI measures improves Alzheimer's disease classification. Hum. Brain Mapp. 37 (5), 1920–1929.

Desikan, R.S., Cabral, H.J., Hess, C.P., Dillon, W.P., Glastonbury, C.M., Weiner, M.W., et al., 2009. Automated MRI measures identify individuals with mild cognitive impairment and Alzheimer's disease. Brain 132 (Pt 8), 2048–2057.

Desikan, R.S., Cabral, H.J., Settecase, F., Hess, C.P., Dillon, W.P., Glastonbury, C.M., et al., 2010. Automated MRI measures predict progression to Alzheimer's disease. Neurobiol. Aging 31 (8), 1364–1374.

Destrieux, C., Fischl, B., Dale, A., Halgren, E., 2010. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. NeuroImage 53 (1), 1–15.

Dukart, J., Schroeter, M.L., Mueller, K., 2011. Alzheimer's Disease Neuroimaging I. Age correction in dementia–matching to a healthy brain. PLoS One 6 (7), e22193.

Eskildsen, S.F., Coupe, P., Garcia-Lorenzo, D., Fonov, V., Pruessner, J.C., Collins, D.L., 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., Ferreira, D., Soininen, H., Mecocci, P., Vellas, B., Tsolaki, M., et al., 2016. The Effect of Age Correction on Multivariate Classification in Alzheimer's Disease, with a Focus on the Characteristics of Incorrectly and Correctly Classified Subjects. Brain

Topogr. 29 (2), 296–307.

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. J. Alzheimers Dis. 41 (3), 685–708.

Fang, R., Wang, G., Huang, Y., Zhuang, J.P., Tang, H.D., Wang, Y., et al., 2014. Validation of the Chinese version of Addenbrooke's cognitive examination-revised for screening mild Alzheimer's disease and mild cognitive impairment. Dement. Geriatr. Cogn. Disord. 37 (3–4), 223–231.

Fazekas, F., Niederkorn, K., Schmidt, R., Offenbacher, H., Horner, S., Bertha, G., et al., 1988. White matter signal abnormalities in normal individuals: correlation with carotid ultrasonography, cerebral blood flow measurements, and cerebrovascular risk factors. Stroke 19 (10), 1285–1288.

Fischl, B., Dale, A.M., 2000. Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proc. Natl. Acad. Sci. U.S.A 97 (20), 11050–11055.

Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., et al., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33 (3), 341–355.

Fischl, B., Sereno, M.I., Dale, A.M., 1999. Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. Neuroimage 9 (2), 195–207.

Gennatas, E.D., Avants, B.B., Wolf, D.H., Satterthwaite, T.D., Ruparel, K., Ciric, R., et al., 2017. Age-Related Effects and Sex Differences in Gray Matter Density, Volume, Mass, and Cortical Thickness from Childhood to Young Adulthood. J. Neurosci. 37 (20), 5065–5073.

Gosche, K.M., Mortimer, J.A., Smith, C.D., Markesbery, W.R., Snowdon, D.A., 2002. Hippocampal volume as an index of Alzheimer neuropathology: findings from the Nun Study. Neurology 58 (10), 1476–1482.

Gouw, A.A., Seewann, A., van der Flier, W.M., Barkhof, F., Rozemuller, A.M., Scheltens, P., et al., 2011. Heterogeneity of small vessel disease: a systematic review of MRI and histopathology correlations. J. Neurol. Neurosurg. Psychiatry 82 (2), 126–135.

Grydeland, H., Westlye, L.T., Walhovd, K.B., Fjell, A.M., 2013. Improved prediction of Alzheimer's disease with longitudinal white matter/gray matter contrast changes. Hum. Brain Mapp. 34 (11), 2775–2785.

Harms, M.P., Somerville, L.H., Ances, B.M., Andersson, J., Barch, D.M., Bastiani, M., et al., 2018. Extending the Human Connectome Project across ages: Imaging protocols for the Lifespan Development and Aging projects. Neuroimage 183, 972–984.

Jack Jr., C.R., Bennett, D.A., Blennow, K., Carrillo, M.C., Dunn, B., Haeberlein, S.B., et al., 2018. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. Alzheimers Dement 14 (4), 535–562.

Jack Jr., C.R., Therneau, T.M., Weigand, S.D., Wiste, H.J., Knopman, D.S., Vemuri, P., et al., 2019. Prevalence of Biologically vs Clinically Defined Alzheimer Spectrum Entities Using the National Institute on Aging-Alzheimer's Association Research Framework. JAMA Neurol., e191971.

Jagust, W.J., Landau, S.M., Koeppe, R.A., Reiman, E.M., Chen, K., Mathis, C.A., et al., 2015. The Alzheimer's Disease Neuroimaging Initiative 2 PET Core: 2015. Alzheimers Dement 11 (7), 757–771.

Jefferson, A.L., Gifford, K.A., Damon, S., Chapman, G.Wt., Liu, D., Sparling, J., et al., 2015. Gray & white matter tissue contrast differentiates Mild Cognitive Impairment converters from non-converters. Brain Imaging Behav. 9 (2), 141–148.

Katzman, R., Zhang, M.Y., Ouang Ya, Q., Wang, Z.Y., Liu, W.T., Yu, E., et al., 1988. A Chinese version of the Mini-Mental State Examination; impact of illiteracy in a Shanghai dementia survey. J. Clin. Epidemiol. 41 (10), 971–978.

Kloppel, S., Stonnington, C.M., Chu, C., Draganski, B., Scahill, R.I., Rohrer, J.D., et al., 2008. Automatic classification of MR scans in Alzheimer's disease. Brain 131 (Pt 3), 681–689.

Kwak, K., Yun, H.J., Park, G., Lee, J.M., 2018. Alzheimer's Disease Neuroimaging I. Multi-Modality Sparse Representation for Alzheimer's Disease Classification. J. Alzheimers Dis. 65 (3), 807–817.

Landau, S.M., Breault, C., Joshi, A.D., Pontecorvo, M., Mathis, C.A., Jagust, W.J., et al., 2013. Amyloid-beta imaging with Pittsburgh compound B and florbetapir: comparing radiotracers and quantification methods. J. Nucl. Med. 54 (1), 70–77.

Landau, S.M., Thomas, B.A., Thurfjell, L., Schmidt, M., Margolin, R., Mintun, M., et al., 2014. Amyloid PET imaging in Alzheimer's disease: a comparison of three radio-tracers. Eur. J. Nucl. Med. Mol. Imaging 41 (7), 1398–1407.

Lindemer, E.R., Greve, D.N., Fischl, B., Augustinack, J.C., Salat, D.H., 2017a. Alzheimer's Disease Neuroimaging I. Differential Regional Distribution of Juxtacortical White Matter Signal Abnormalities in Aging and Alzheimer's Disease. J. Alzheimers Dis. 57 (1), 293–303.

Lindemer, E.R., Greve, D.N., Fischl, B., Salat, D.H., Gomez-Isla, T., 2018. White matter abnormalities and cognition in patients with conflicting diagnoses and CSF profiles. Neurology 90 (17), e1461–e1469.

Lindemer, E.R., Greve, D.N., Fischl, B.R., Augustinack, J.C., Salat, D.H., 2017b. Regional staging of white matter signal abnormalities in aging and Alzheimer's disease. Neuroimage Clin. 14, 156–165.

Lindemer, E.R., Salat, D.H., Smith, E.E., Nguyen, K., Fischl, B., Greve, D.N., et al., 2015. White matter signal abnormality quality differentiates mild cognitive impairment that converts to Alzheimer's disease from nonconverters. Neurobiol. Aging 36 (9), 2447–2457.

Mohs, R.C., Knopman, D., Petersen, R.C., Ferris, S.H., Ernesto, C., Grundman, M., et al., 1997. Development of cognitive instruments for use in clinical trials of antidementia drugs: additions to the Alzheimer's Disease Assessment Scale that broaden its scope. The Alzheimer's Disease Cooperative Study. Alzheimer Dis. Assoc. Disord. 11 (Suppl 2), S13–S21.

Nasrabady, S.E., Rizvi, B., Goldman, J.E., Brickman, A.M., 2018. White matter changes in Alzheimer's disease: a focus on myelin and oligodendrocytes. Acta Neuropathol. Commun. 6 (1), 22.

Ossenkoppele, R., Smith, R., Ohlsson, T., Strandberg, O., Mattsson, N., Insel, P.S., et al.,

2019. Associations between tau, Abeta, and cortical thickness with cognition in Alzheimer disease. Neurology 92 (6), e601–e612.

Park, M., Moon, W.J., 2016. Structural MR Imaging in the Diagnosis of Alzheimer's Disease and Other Neurodegenerative Dementia: Current Imaging Approach and Future Perspectives. Korean J Radiol 17 (6), 827–845.

Polat, K., Güneş, S., 2009. A new feature selection method on classification of medical datasets: Kernel F-score feature selection. Expert Syst. Appl. 36 (7), 10367–10373.

Raamana, P.R., Weiner, M.W., Wang, L., Beg, M.F., 2015. Alzheimer's Disease Neuroimaging I. Thickness network features for prognostic applications in dementia. Neurobiol. Aging 36 (Suppl 1), S91–S102.

Salat, D.H., 2014. Imaging small vessel-associated white matter changes in aging. Neuroscience 276, 174–186.

Salat, D.H., Chen, J.J., van der Kouwe, A.J., Greve, D.N., Fischl, B., Rosas, H.D., 2011. Hippocampal degeneration is associated with temporal and limbic gray matter/white matter tissue contrast in Alzheimer's disease. Neuroimage 54 (3), 1795–1802.

Salat, D.H., Lee, S.Y., van der Kouwe, A.J., Greve, D.N., Fischl, B., Rosas, H.D., 2009. Age-associated alterations in cortical gray and white matter signal intensity and gray to white matter contrast. Neuroimage 48 (1), 21–28.

Salvatore, C., Battista, P., Castiglioni, I., 2016. Frontiers for the Early Diagnosis of AD by Means of MRI Brain Imaging and Support Vector Machines. Curr. Alzheimer Res. 13 (5), 509–533.

Schwarz, C.G., Gunter, J.L., Wiste, H.J., Przybelski, S.A., Weigand, S.D., Ward, C.P., et al., 2016. A large-scale comparison of cortical thickness and volume methods for measuring Alzheimer's disease severity. Neuroimage Clin. 11, 802–812.

Sengoku, R., 2019. Aging and Alzheimer's disease pathology. Neuropathology.

Smith, S.M., Vidaurre, D., Alfaro-Almagro, F., Nichols, T.E., Miller, K.L., 2019. Estimation of brain age delta from brain imaging. Neuroimage 200, 528–539.

Sowell, E.R., Peterson, B.S., Kan, E., Woods, R.P., Yoshii, J., Bansal, R., et al., 2007. Sex differences in cortical thickness mapped in 176 healthy individuals between 7 and 87 years of age. Cereb. Cortex 17 (7), 1550–1560.

Szabo, C.A., Lancaster, J.L., Xiong, J., Cook, C., Fox, P., 2003. MR imaging volumetry of subcortical structures and cerebellar hemispheres in normal persons. AJNR Am. J. Neuroradiol. 24 (4), 644–647.

Tang, X., Holland, D., Dale, A.M., Younes, L., Miller, M.I., 2014. Alzheimer's Disease Neuroimaging I. Shape abnormalities of subcortical and ventricular structures in mild cognitive impairment and Alzheimer's disease: detecting, quantifying, and predicting. Hum. Brain Mapp. 35 (8), 3701–3725.

Teipel, S.J., Grothe, M., Lista, S., Toschi, N., Garaci, F.G., Hampel, H., 2013. Relevance of magnetic resonance imaging for early detection and diagnosis of Alzheimer disease. Med. Clin. North Am. 97 (3), 399–424.

Tentolouris-Piperas, V., Ryan, N.S., Thomas, D.L., Kinnunen, K.M., 2017. Brain imaging evidence of early involvement of subcortical regions in familial and sporadic Alzheimer's disease. Brain Res. 1655, 23–32.

Varatharajah, Y., Ramanan, V.K., Iyer, R., Vemuri, P., 2019. Alzheimer's Disease Neuroimaging I. Predicting Short-term MCI-to-AD Progression Using Imaging, CSF, Genetic Factors, Cognitive Resilience, and Demographics. Sci Rep. 9 (1), 2235.

Wachinger, C., Salat, D.H., Weiner, M., Reuter, M., 2016. Alzheimer's Disease Neuroimaging I. Whole-brain analysis reveals increased neuroanatomical asymmetries in dementia for hippocampus and amygdala. Brain 139 (Pt 12), 3253–3266.

Westlye, L.T., Walhovd, K.B., Dale, A.M., Espeseth, T., Reinvang, I., Raz, N., et al., 2009. Increased sensitivity to effects of normal aging and Alzheimer's disease on cortical thickness by adjustment for local variability in gray/white contrast: a multi-sample MRI study. Neuroimage 47 (4), 1545–1557.

Yao, Z., Hu, B., Liang, C., Zhao, L., Jackson, M., 2012. Alzheimer's Disease Neuroimaging I. A longitudinal study of atrophy in amnestic mild cognitive impairment and normal aging revealed by cortical thickness. PLoS One 7 (11), e48973.