Received: 23 June 2017 | Revised: 28 November 2017 | Accepted: 7 December 2017

DOI: 10.1002/hbm.23922

RESEARCH ARTICLE

WILEY

Factors influencing accuracy of cortical thickness in the diagnosis of Alzheimer's disease

Mahanand Belathur Suresh[1,2,3] | Bruce Fischl[1,2,4] | David H. Salat[1,2,5] | for the Alzheimer's Disease Neuroimaging Initiative (ADNI)*

[1]MGH/MIT/HMS Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School, Charlestown, Massachusetts

[2]Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts

[3]Department of Information Science and Engineering, Sri Jayachamarajendra College of Engineering, Mysuru, Karnataka, India

[4]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts

[5]Neuroimaging Research for Veterans Center, VA Boston Healthcare System, Boston, Massachusetts

Correspondence
Mahanand Belathur Suresh; Department of Information Science and Engineering, Sri Jayachamarajendra College of Engineering, Mysuru 570 006, Karnataka, India. Email; bsmahanand@sjce.ac.in

Funding Information
National Institutes of Health - NIH, Grant/Award Numbers: R01NR010827, NS042861, NS058793; Center for Functional Neuroimaging Technologies, Grant/Award Number: P41RR14075; Biomedical Technology Program of the National Center for Research Resources (NCRR), NIH; NCRR Shared Instrumentation Grant Program; High-End Instrumentation Grant Program, Grant/Award Numbers: S10RR021110, S10RR023401, S10RR019307, S10RR019254, S10RR023043; University Grants Commission, Government of India; Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health, Grant/Award Number: U01 AG024904; DOD ADNI (Department of Defense, Grant/Award Number: W81XWH-12-2-0012; National Institute on Aging; National Institute of Biomedical Imaging and Bioengineering; Canadian Institutes of Health Research

Abstract

There is great value to use of structural neuroimaging in the assessment of Alzheimer's disease (AD). However, to date, predictive value of structural imaging tend to range between 80% and 90% in accuracy and it is unclear why this is the case given that structural imaging should parallel the pathologic processes of AD. There is a possibility that clinical misdiagnosis relative to the gold standard pathologic diagnosis and/or additional brain pathologies are confounding factors contributing to reduced structural imaging classification accuracy. We examined potential factors contributing to misclassification of individuals with clinically diagnosed AD purely from cortical thickness measures. Correctly classified and incorrectly classified groups were compared across a range of demographic, biological, and neuropsychological data including cerebrospinal fluid biomarkers, amyloid imaging, white matter hyperintensity (WMH) volume, cognitive, and genetic factors. Individual subject analyses suggested that at least a portion of the control individuals misclassified as AD from structural imaging additionally harbor substantial AD biomarker pathology and risk, yet are relatively resistant to cognitive symptoms, likely due to "cognitive reserve," and therefore clinically unimpaired. In contrast, certain clinical control individuals misclassified as AD from cortical thickness had increased WMH volume relative to other controls in the sample, suggesting that vascular conditions may contribute to classification accuracy from cortical thickness measures. These results provide examples of factors that contribute to the accuracy of structural imaging in predicting a clinical diagnosis of AD, and provide important information about considerations for future work aimed at optimizing structural based diagnostic classifiers for AD.

KEYWORDS

Alzheimer's disease, cortical thickness, magnetic resonance imaging, support vector machines, white matter hyperintensity

1 | INTRODUCTION

Alzheimer's Disease (AD) is the most common form of dementia in older adults and is characterized by significant loss or decline in memory, problems in learning, and other cognitive abilities (Bäckman, Jones, Berger, Laukka, & Small, 2004; Burns and Iliffe, 2009; Carlesimo and

*Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (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.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Oscar-Berman, 1992; Dubois et al., 2014; Holtzman, Morris, & Goate, 2011; McKhann et al., 1984, 2011; Querfurth and LaFerla, 2010). The current gold standard for a conclusive diagnosis of AD is through post-mortem examination to identify the disease-defining regional patterns of neurofibrillary tau tangles and amyloid plaque deformities in the brain (Arnold, Hyman, Flory, Damasio, & Van Hoesen, 1991; Braak and Braak, 1991; Brun and Gustafson, 1976; Hardy, 2006; Hyman et al., 2012; Montine et al., 2012; Selkoe and Hardy, 2016). Mild cognitive impairment (MCI) is considered as a potential intermediate stage between normal aging and AD. Individuals with an MCI diagnosis have an increased risk of developing AD and conversion rate is reported to be approximately 10%–15% per year (Petersen et al., 2001). Although MCI is known to be a clinically heterogeneous condition, clinical diagnosis of probable AD is assumed to be a more stable condition that can be achieved with reasonable accuracy relative to the pathologic gold standard (Joachim, Morris, & Selkoe, 1988; Lopez et al., 2000). However, clinical misdiagnosis relative to pathology can be substantial and depends on several factors. For example, a recent review across a total of 919 samples with at least one clinical visit and autopsy demonstrated that clinical diagnosis of probable and possible AD has a sensitivity ranged from 70.9% to 87.3% while specificity ranged from 44.3% to 70.8% relative to the pathological diagnosis (Beach, Monsell, Phillips, & Kukull, 2012). The implication of this is that there is a relative uncertainty in the clinical diagnosis of AD as compared to the neuropathological diagnosis. Additionally, clinical diagnoses are often achieved late in the disease process when substantial brain tissue damage has resulted in noticeable cognitive deficit and other symptoms. Early clinical diagnosis is likely to have greater inaccuracy relative to the neuropathologic criteria (which is why diagnosis of MCI due to AD in the absence of biomarkers is particularly difficult). Given potential for clinical misdiagnosis and the need for early diagnostics, great emphasis has been put towards neuroimaging approaches for robust diagnosis of AD, prior to symptom development.

Structural imaging procedures have been successful in AD diagnostics (Chetelat and Baron, 2003; Frisoni, Fox, Jack, Scheltens, & Thompson, 2010; Park and Moon, 2016; Teipel et al., 2013). This is due to the fact that structural atrophy mirrors patterns of the characteristic regional pathology of this condition (Jack, Petersen, O'brien, & Tangalos, 1992; Jack et al., 1997; Scheltens et al., 1992). Several prior studies have applied structural imaging procedures in the later as well as earlier, and "preclinical" (e.g., MCI) stages of the disease (Dickerson et al., 2009; Eskildsen et al., 2013; Fotenos, Snyder, Girton, Morris, & Buckner, 2005; Jack et al., 1999; Killiany et al., 2000, 2002; Scheltens, Fox, Barkhof, & De Carli, 2002). Given current goals of clinical trials for identifying individuals in the earliest stages of disease, structural imaging could be advantageous as an initial screen for molecular procedures such as positron emission tomography which is currently used but costly and invasive.

AD detection using machine learning and magnetic resonance imaging (MRI) is a promising area of research. Support vector machines (SVM) is the most widely used procedure in neuroimaging studies to classify AD (Aguilar et al., 2013; Cuingnet et al., 2011; Klöppel et al., 2008; Magnin et al., 2009; Salvatore, Battista, & Castiglioni, 2016; Schmitter et al., 2015; Vemuri et al., 2008; Wolz et al., 2011). Although relatively successful, review of prior work demonstrates that there seems to be a "hard" limitation to accuracy with most studies reporting a range between 80 and 90%. There are two likely explanations for this limitation. First, to our knowledge, no prior work has been performed using a gold standard pathologic diagnosis and therefore all studies will have some degree of clinical misdiagnosis that contributes detrimentally to both the training and testing phases of classification. Second, it is likely that biological variability due in part to aging and other health and disease factors contribute to each individual's pattern of cortical thinning in a manner that may confound classification. To our knowledge, no prior studies have examined how these many factors contribute to machine learning classification accuracy of AD from structural MRI, however, recent studies have uncovered the need to consider age as a confounding factor contributing to misclassification of individuals with AD (Dukart et al., 2011; Falahati et al., 2016). Little is currently known however about misclassification of AD from structural MRI when age is removed as a factor.

Current strategies for therapeutic clinical trials aim to identify individuals in the earliest stages of AD. Such a task should be greatly enhanced by structural imaging, yet it is critical to know factors that influence classification based on brain structure alone. Although it is possible to enhance classification through overfitting (e.g., including a range of variables in the classifier that are not thought to be directly related to AD pathology), this would in fact decrease the pathologic accuracy of the classification by including individuals with impairment due to other etiologies which would confound therapeutic trials aimed specifically at AD pathology. The goal of our work was to define factors that contribute to the known 10%–20% misclassification from structural MRI, as opposed to achieving the greatest possible accuracy in classification. Specifically, we aimed to identify demographic and biological factors that were most influential on the accuracy of classification of individuals with a clinical diagnosis of probable AD exclusively using regional cortical atrophy patterns given known links between cortical atrophy and regional patterns of AD histopathology. This work follows on recent efforts to optimally integrate biomarker information into the early and accurate diagnosis of individuals with pathologic AD (Falahati et al., 2016; Hwang et al., 2016; Jack et al., 2016; Landau et al., 2010; Mattsson et al., 2015; Palmqvist et al., 2015).

Structural imaging was achieved with measurements of cortical thickness based on MRI which has been shown to index pathology in AD (Bakkour, Morris, & Dickerson, 2009; Dickerson et al., 2009; Lerch et al., 2008; McDonald et al., 2009; Salat et al., 2011). Cortical thickness measurements are sensitive to subtle degenerative changes making this imaging marker an ideal feature for AD classification (de Vos et al., 2016; Eskildsen et al., 2013; Raamana et al., 2015; Wolz et al., 2011). Classification performance was examined relative to demographic factors and biomarkers in misclassified individuals to gain insight into potential causes of misclassification including age, sex, education, Mini Mental State Examination (MMSE), cerebrospinal fluid (CSF) tau and amyloid-beta, Florbetapir amyloid PET positivity (F-AV45), APOE4 genotype, scanner type, American National Adult Reading Test (ANART) score, and Rey Auditory Verbal Learning Test

TABLE 1 Subject demographics from the ADNI dataset

	Entire study		Cortical thickness study	
	Control	AD	Control	AD
Number of subjects	269	137	50	50
Sex (Female/Male)	149/120	58/79	25/25	25/25
Age, years (mean)	72.86	74.16	72.81	72.77
Education, years (mean)	16.58	15.79	16.98	16.58

(RAVLT) scores. We additionally examined the influence of white matter hyperintensity (WMH) burden on classification given this common type of brain tissue alteration in both typical aging as well as AD. WMH are highly prevalent in older adults and in AD and are related to brain structural measures, and therefore may be an important factor related to classification accuracy (de Leeuw, 2001; Debette and Markus, 2010; Hopkins et al., 2006; Mortamais et al., 2013; Murray et al., 2005; Provenzano et al., 2013; Ylikoski et al., 1995). Such information could be useful in determining whether misclassification was due to potential biological variability, clinical misdiagnosis, and/or technical limitations of the procedures and this information could be used toward further improvement of structural imaging procedures for the diagnosis of AD and provide potential insight into mechanisms of variability in cognitive expression of AD.

2 | MATERIALS AND METHODS

2.1 | Dataset

We used the structural brain MRI scans from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset (adni.loni.usc.edu). The ADNI was launched in 2003 as a public private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging, positron emission tomography, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment and early AD. For up-to-date information, see www.adni-info.org. A total of 406 subjects (269 controls and 137 AD) were considered in this study. Standard 3 T baseline T1-weighted images were included from the ADNI data set.

Several studies have reported the effect of aging on global and regional brain changes in controls and AD (Lim, Zipursky, Murphy, & Pfefferbaum, 1990; Salat, Kaye, & Janowsky, 1999; Salat, Kaye, & Janowsky, 2001; Salat et al., 2004; Shear et al., 1995; Thompson et al., 1998). As noted previously, age-related brain changes can potentially lead to misclassification of younger AD patients and older control individuals. We therefore created matched control and patient groups for age and other demographic factors. 100 subjects (50 Controls and 50 AD) matched for age, sex, and education were used in the surface based cortical thickness analysis and the remaining 306 subjects were used for classification. Subject demographics for the entire study and for the matched groups used for cortical thickness study are presented in Table 1.

2.2 | Cortical thickness measurement

The FreeSurfer image analysis suite version 5.3.0 (http://surfer.nmr.mgh.harvard.edu) was employed to process the MRI data and compute cortical thickness measurements. The technical details of cortical reconstruction and volumetric segmentation performed with the Freesurfer are described in prior publications (Dale, Fischl, & Sereno, 1999; Dale and Sereno, 1993; Fischl, Sereno, & Dale, 1999a; Fischl, Sereno, Tootell, & Dale, 1999b; Fischl and Dale, 2000; Fischl, Liu, & Dale, 2001; Fischl et al., 2002, 2004a, 2004b; Han et al., 2006; Jovicich et al., 2006; Reuter, Rosas, & Fischl, 2010; Reuter, Schmansky, Rosas, & Fischl, 2012; Segonne et al., 2004). Briefly, this processing includes motion correction and averaging of multiple volumetric T1 weighted images (M. Reuter et al., 2010), removal of nonbrain tissue using a hybrid watershed/surface deformation procedure (Segonne et al., 2004), automated Talairach transformation, segmentation of the subcortical white matter and deep gray matter volumetric structures (Fischl et al., 2002, 2004a), intensity normalization (Sled, Zijdenbos, & Evans, 1998), tessellation of the gray matter white matter boundary, automated topology correction (Fischl et al., 2001; Segonne, Pacheco, & Fischl, 2007), and surface deformation following intensity gradients to optimally place the gray/white and gray/cerebrospinal fluid borders at the location where the greatest shift in intensity defines the transition to the other tissue class (Dale et al., 1999; Dale and Sereno, 1993; Fischl and Dale, 2000). Once the cortical models are complete, a number of deformable procedures can be performed for further data processing and analysis including surface inflation (Fischl et al., 1999a, 1999b), registration to a spherical atlas which is based on individual cortical folding patterns to match cortical geometry across subjects (Fischl et al., 1999a, 1999b), parcellation of the cerebral cortex into units with respect to gyral and sulcal structure (Desikan et al., 2006; Fischl et al., 2004a), and creation of a variety of surface based data including maps of curvature and sulcal depth. This method uses both intensity and continuity information from the entire three dimensional MR volume in segmentation and deformation procedures to produce representations of cortical thickness, calculated as the closest distance from the gray/white boundary to the gray/CSF boundary at each vertex on the tessellated surface (Fischl and Dale, 2000). The maps are created using spatial intensity gradients across tissue classes and are therefore not simply reliant on absolute signal intensity. The maps produced are not restricted to the voxel resolution of the original data thus are capable of detecting submillimeter differences between groups. Procedures for the measurement of cortical thickness have been validated against histological analysis (Rosas et al., 2002) and manual measurements (Kuperberg et al., 2003; Salat et al., 2004). Thickness measurements were mapped on the inflated surface of each participant's reconstructed brain (Dale et al., 1999; Fischl et al., 1999a, 1999b). This procedure allowed visualization of data across the entire cortical surface (i.e., both the gyri and sulci) without interference from cortical folding. Maps were subsequently smoothed using a circularly symmetric Gaussian kernel across the surface with a full-width-half-maximum (FWHM) of 10 mm. Next, cortical maps were averaged across participants using a nonrigid high-dimensional spherical

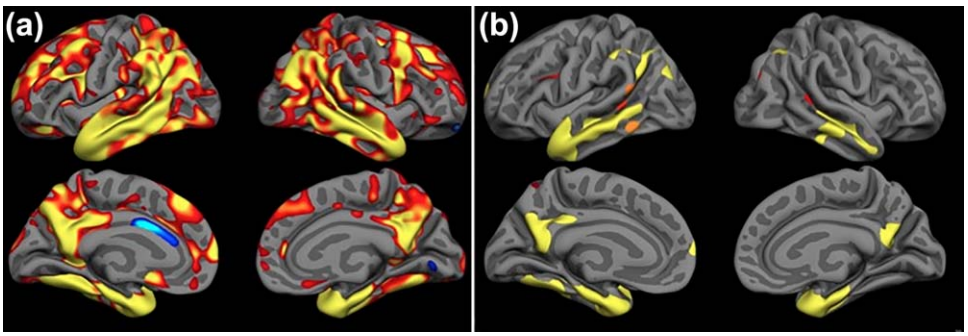

FIGURE 1 (a) Surface maps of the cortical thickness differences between the controls and AD groups smoothed on the surface with an approximate Gaussian kernel of a full-width-half-max (FWHM) of 10 mm ($p < .05$ uncorrected). (b) Surface maps of the cortical thickness differences after the correction for multiple comparisons (thresholded at $p < .0001$) [Color figure can be viewed at wileyonlinelibrary.com]

averaging method to align cortical folding patterns (Fischl et al., 1999a, 1999b). This procedure provides accurate matching of morphologically homologous cortical locations among participants on the basis of each individual's anatomy while minimizing metric distortion, resulting in a mean measure of cortical thickness at each point on the reconstructed surface.

2.3 | Statistical analysis

Classifier features were selected as regions showing statistical thinning in the AD group compared to the control group. Whole-brain surface-based general linear models were performed at each surface vertex (10,242 vertices per hemisphere). Resulting *z*-statistic maps were thresholded at $p < .05$. Multiple comparison correction was then performed using a clusterwise procedure (Hagler, Saygin, & Sereno, 2006). Surface data were corrected for multiple comparisons with a threshold of $p < .0001$ for both left and right hemispheres.

2.4 | Support vector machine classifier

Support vector machine is a commonly utilized supervised, multivariate classification method. The problem of AD detection using SVM was formulated as a binary classification problem. In brief, given an *N* observation samples $\{\boldsymbol{x}_i, y_i\}$, where $\boldsymbol{x}_i = [x_{i1}, \ldots, x_{in}] \in \Re^n$ and $y_i \in [1, -1]$ is the coded group label. If the sample belongs to that of an AD patient then the coded group label y_i is defined as one ($y_i = 1$), otherwise, sample belongs to control where the coded group label is ($y_i = -1$). The SVM classifier finds a hyperplane maximizing the margin between groups. More details on SVM can be found in (Cortes and Vapnik, 1995; Schölkopf and Smola, 2002). In this study, we used the SVM implementation publicly available in LibSVM (csie.ntu.edu.tw/~cjlin/libsvm). To determine a suitable cost parameter *C* and kernel parameter γ of the nonlinear Gaussian function in the SVM classifier, the parameter values are optimized using a cross-validation via grid-search approach (Chang and Lin, 2011). The grid search was performed over the ranges $C = 2^{-5}, 2^{-2}, \ldots, 2^{15}$, $\gamma = 2^{-15}, 2^{-13}, \ldots, 2^{5}$. The optimized set of parameters was then used to train the SVM classifier using the training set. The cortical thickness values obtained using surface based thickness analysis were used as features in the SVM and classifier was trained to predict the group membership on the remaining 306 subjects. A total of 22 features (cortical thickness regions) were used for classification. We used the 10-fold cross-validation to estimate the SVM classifier performance. During each fold, the classifier was developed using 90% of the subjects as training data and remaining 10% of the subjects as testing data. For better generalization, the classifier was tested for 100 different random combination of training and testing datasets.

3 | RESULTS

3.1 | Cortical thickness comparison between controls and AD

Cortical thickness measurements were obtained using FreeSurfer on 100 subjects (50 Controls and 50 AD) matched for age, gender, and education. As expected, group differences in cortical thickness between the controls and AD were robust and are demonstrated in Figure 1.

Regions exhibiting significantly reduced cortical thickness in AD compared to controls following cluster-based multiple comparison correction based on the Freesurfer Desikan/Killiany parcellation atlas (Desikan et al., 2006) were 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. Mean cortical thickness within each significant region was extracted from the remaining 306 subjects, and used as an input to the SVM classifier.

3.2 | Performance of SVM classifier

To illustrate the performance of the SVM classifier, mean cortical thickness values for each of the significant regions of interest were extracted from the remaining 306 subjects and used as features. A total of 22 features (cortical thickness regions) were used for classification. We used the 10-fold cross-validation to obtain an unbiased estimate of the classifier performance. During each fold the classifier was

developed using data from 90% of the subjects and tested using data from the remaining 10% of the subjects. For better generalization, the classifier was tested for 100 different random combination of training and testing datasets. 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 our experiments, mean accuracy of 90.32% (standard deviation 4.02), mean sensitivity of 0.96 (standard deviation 0.03), and mean specificity of 0.76 (standard deviation 0.14) were obtained using the SVM classifier.

Hippocampal volume has been used in several prior classification studies and is very effective to achieving classification to similar degrees as other structural measures. We therefore performed a secondary classifier analysis incorporating this imaging information with the initial classifier. Inclusion of hippocampal volume (corrected for ICV) as a feature increased the accuracy by 2.43%; however, we did not include hippocampal volume as a feature in the final classifier. This is because the goal of the work was to utilize structural features that are believed to be most related to the primary pathologic features in AD for a greater potential ability to identify individuals with true AD pathology. Our prior work demonstrated that hippocampal volume statistically factors with white matter lesions (WMH) which are not considered a primary pathologic feature of AD (Coutu et al., 2016, 2017). Inclusion of hippocampal volume in the classifier may slightly improve accuracy with regard to clinical diagnosis, but this may not be concordant with the pathologic diagnosis of AD. We therefore did not include hippocampal volume to attempt to minimize vascular related variance in the classification.

We additionally performed a classification analysis using whole-cortex surface measurements. Cortical thickness values from the whole-cortex were used as features to train an SVM classifier with an accuracy of 89.27% (standard deviation 3.49), sensitivity of 0.97 (standard deviation 0.02), and specificity of 0.74 (standard deviation 0.08). Although in the high-performance range, results indicate that cortical thickness could not perfectly classify clinical group membership using the SVM classifier.

3.3 | Analysis to determine factors contributing to misclassification

The results obtained using SVM classifier demonstrate that ~10% of the subjects are misclassified based on structural imaging compared to their clinical diagnosis based on thickness values alone. These findings show that the procedures are robust for a large majority of the sample and therefore additional information about failures would better inform future implementations of SVM. Follow-up analyses examined potential factors that may contribute to any individual being misclassified. As noted, limitations in accuracy are potentially related to technical restrictions, incorrect clinical diagnosis or the confounding effects of demographic and/or health disparities rather than the method used for classification (Falahati, Westman, & Simmons, 2014). In the following analyses, we therefore compared variables of interest across four groups: controls correctly classified as controls (classified CN), controls classified as AD (misclassified CN), AD correctly classified as AD (classified AD), and AD classified as controls (misclassified AD).

Demographic information. Initial analysis demonstrated that groups differed with regard to age as expected from prior work. Specifically, older control individuals were more likely to be classified as AD. However, age was not the only factor contributing to misclassification. We therefore created age and sex matched subsets of the four groups classified CN, misclassified CN, classified AD, and misclassified AD for subsequent comparison.

Cortical thickness. Given that the SVM is based on regional patterns of cortical thinning, it was expected that groups classified as having AD would have generally thinner cortex than groups classified as control. However, it is unknown whether these results would be regionally selective or more global. Thus, although a somewhat circular analysis (given that the classifier was based on thickness values) we next compared age and sex matched groups using surface based thickness general linear models to better understand the spatial nature of the thickness patterns in the misclassified groups for illustrative purposes.

Surface maps of the cortical thickness differences between correctly classified controls and misclassified controls revealed reduced cortical thickness in misclassified controls in several cortical regions (Figure 2a) including inferior parietal cortex, superior frontal gyrus, medial orbital frontal cortex, inferior parietal cortex, superior temporal gyrus, superior frontal gyrus, superior parietal cortex, entorhinal cortex, precuneus, postcentral gyrus, middle temporal gyrus, pars orbitalis, precentral gyrus, and lateral occipital cortex. In contrast, thickness differences between correctly classified AD and misclassified AD revealed that correctly classified AD had significantly reduced thickness in several regions including overlap with most of the classifier regions. Thicker cortex in misclassified AD was found in supramarginal gyrus, middle temporal gyrus, insula, inferior temporal gyrus, precentral gyrus, lateral orbital frontal cortex, superior frontal gyrus, precentral gyrus, superior parietal cortex, lingual gyrus, inferior parietal cortex, pars triangularis, entorhinal cortex, middle temporal gyrus, lateral occipital cortex, fusiform gyrus, and pars orbitalis (Figure 2b).

Cortical thickness comparison between misclassified control and correctly classified AD demonstrated that although the misclassified controls had reduced cortical thickness relative to correctly classified controls (Figure 2a), they still had significantly thicker cortex than correctly classified AD in typical AD pathology regions. The regions included entorhinal cortex, middle temporal gyrus, fusiform gyrus, inferior temporal gyrus, supramarginal gyrus, middle temporal gyrus, inferior parietal cortex, and parahippocampal gyrus (Figure 2c) suggesting that any pathology in those regions may be an earlier stage and/or a different pathophysiological process. The misclassified controls have thinning in classifiers regions (regions sensitive to AD pathology). Compared to correctly classified controls, it would seem that this is a more global degenerative effect, and compared to correctly classified AD, the AD sensitive regions are relatively more affected in AD compared to the misclassified controls. This suggests that addition of features of

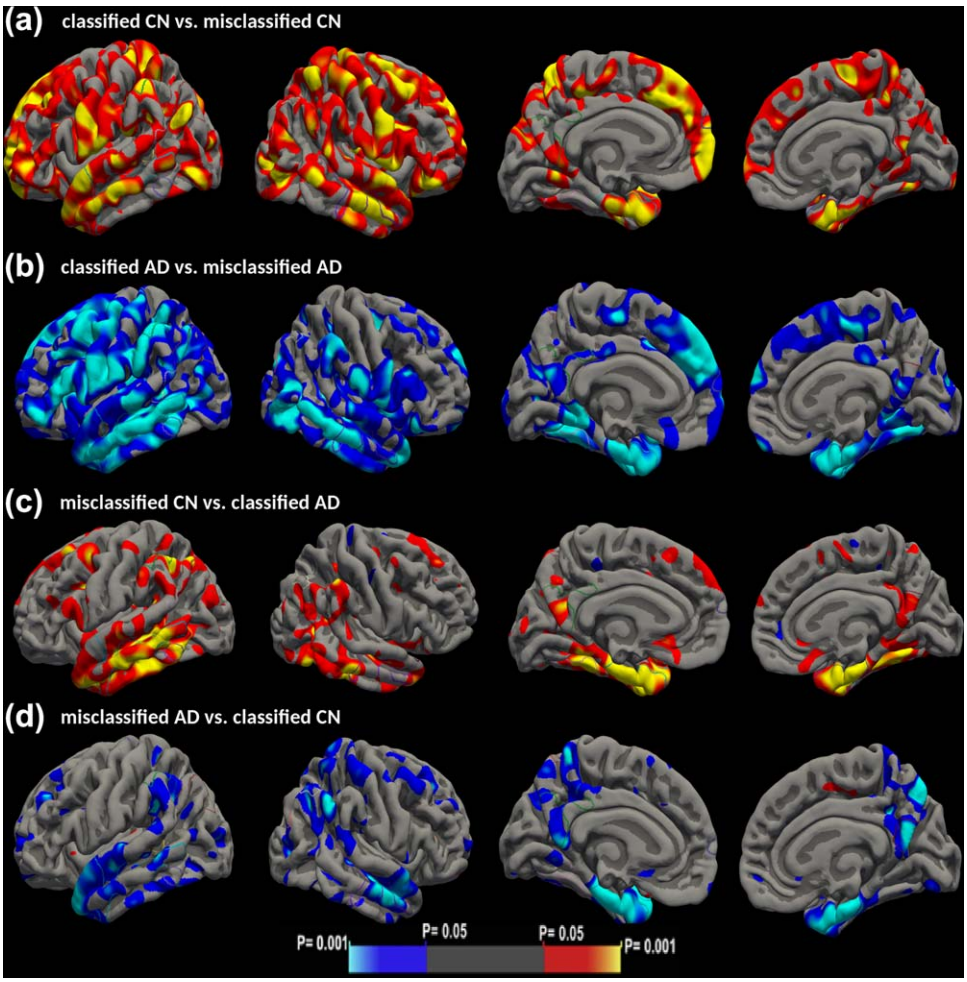

FIGURE 2 Surface maps of the cortical thickness differences between the classified and misclassified groups smoothed on the surface with an approximate Gaussian kernel of an FWHM of 10 mm [Color figure can be viewed at wileyonlinelibrary.com]

the proportion of change in AD regions compared to other non-AD-specific regions could be an important addition to the classification process. Misclassified AD compared to correctly classified controls, exhibited greater cortical thickness in selective regions overlapping the classifier regions (Figure 2d). The regions including parahippocampal gyrus, superior temporal gyrus, inferior parietal cortex, entorhinal cortex, middle temporal gyrus, precuneus cortex, and supramarginal gyrus potentially suggest an earlier stage of disease pathology. The misclassified AD compared to controls has a pattern of thinning that is similar to the AD classifier regions. In theory, these individuals should have been detected by the classifier. It is possible that these individuals are therefore in an earlier stage of pathology and may require additional features to be accurately classified.

We next examined a series of variables that are associated with a diagnosis or enhanced risk for Alzheimer's disease including MMSE score, CSF tau and amyloid-beta, (18) F-AV45, APOE4 genotype, scanner type, and WMH volume. We used WMH available through the ADNI database, calculated from fluid attenuated inversion recovery (FLAIR) and T1-weighted images using a Bayesian segmentation method (DeCarli, Murphy, Teichberg, Campbell, & Sobering, 1996; DeCarli et al., 1999; DeCarli, Pauline, & Evan, 2013). The WMH volume was expressed as a portion of intracrainial volume (WMH/ICV $\times$ 100) to correct for head size. CSF tau was used to pick extreme subjects that were unlikely to have a pathologic diagnosis matching the clinical diagnosis to illustrate how the factors of interest contribute to clinical classification accuracy. A single misclassified control subject (ADNI subject Id: 127_S_5185) was selected based on having the greatest CSF tau levels in that group (levels suggestive of AD pathology) represented as red circle and a single misclassified AD subject (ADNI subject Id: 052_S_4959) was selected based on having the lowest CSF tau levels in that group (levels suggestive of lack of tau pathology) represented a green circle for example comparisons. These single subjects are not presented as representative of the group, but simply as examples of the individual variability within the misclassified group. Specifically, we aimed to examine in detail example individuals that were most likely to have biomarker data that was inconsistent with their clinical diagnosis as potentially salient examples of factors influencing structural imaging classifier accuracy. Plots showing each of these markers across the four groups are demonstrated in Figure 3.

We also examined educational level, ANART score, and RAVLT scores to assess the verbal learning, memory performance, and premorbid verbal intelligence of the groups. Plots showing each of these factors across the four groups are demonstrated in Figure 4.

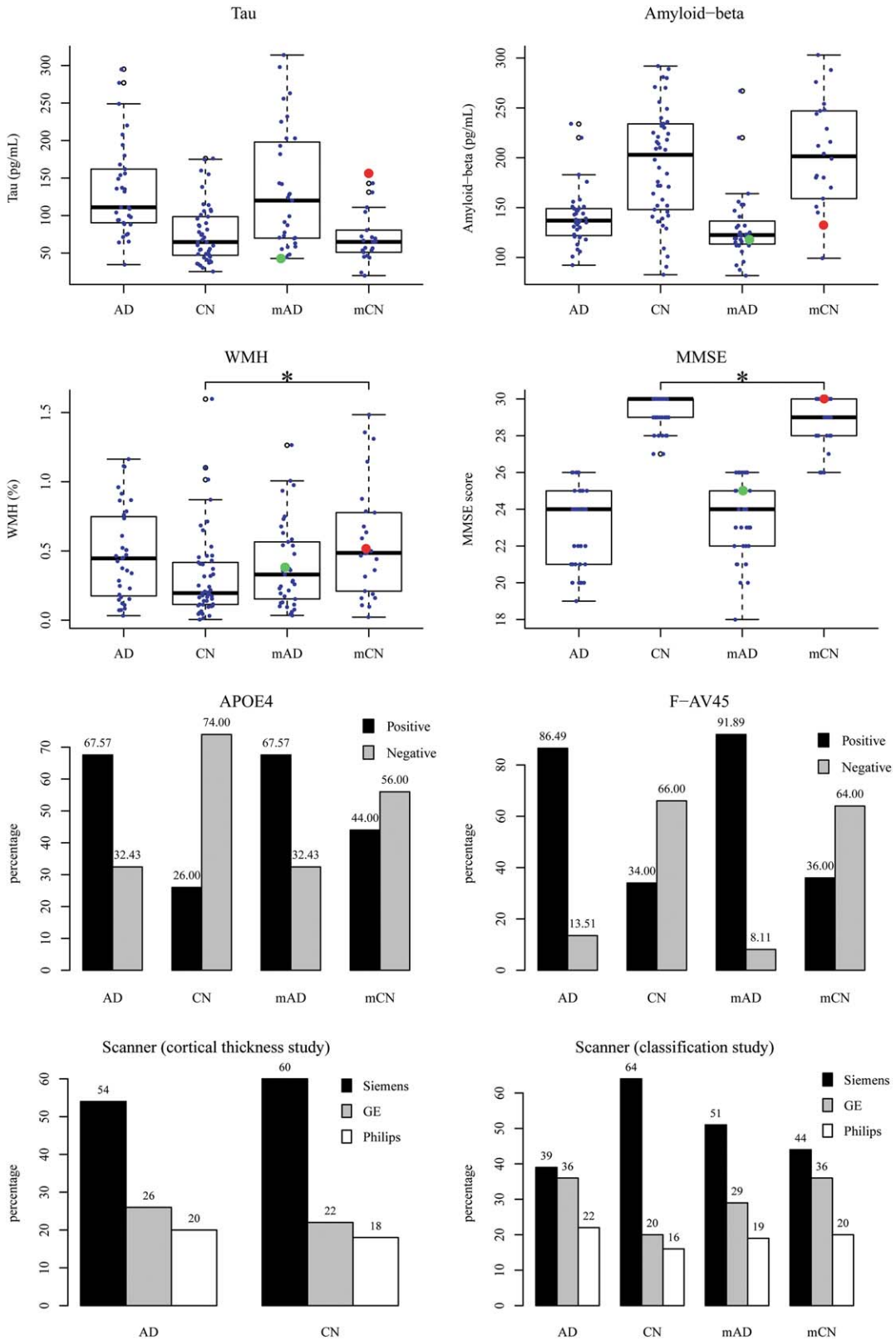

FIGURE 3 Plots showing the distribution of CSF biomarkers, amyloid imaging, white matter hyperintensity, cognitive, genetic factors, and scanner types. A single misclassified control subject (ADNI subject Id: 127_S_5185) is represented as red circle and a single misclassified AD subject (ADNI subject Id: 052_S_4959) is represented as green circle. The symbol * indicates significantly different. AD = classified AD, mAD = misclassified AD, CN = classified control, mCN = misclassified control [Color figure can be viewed at wileyonlinelibrary.com]

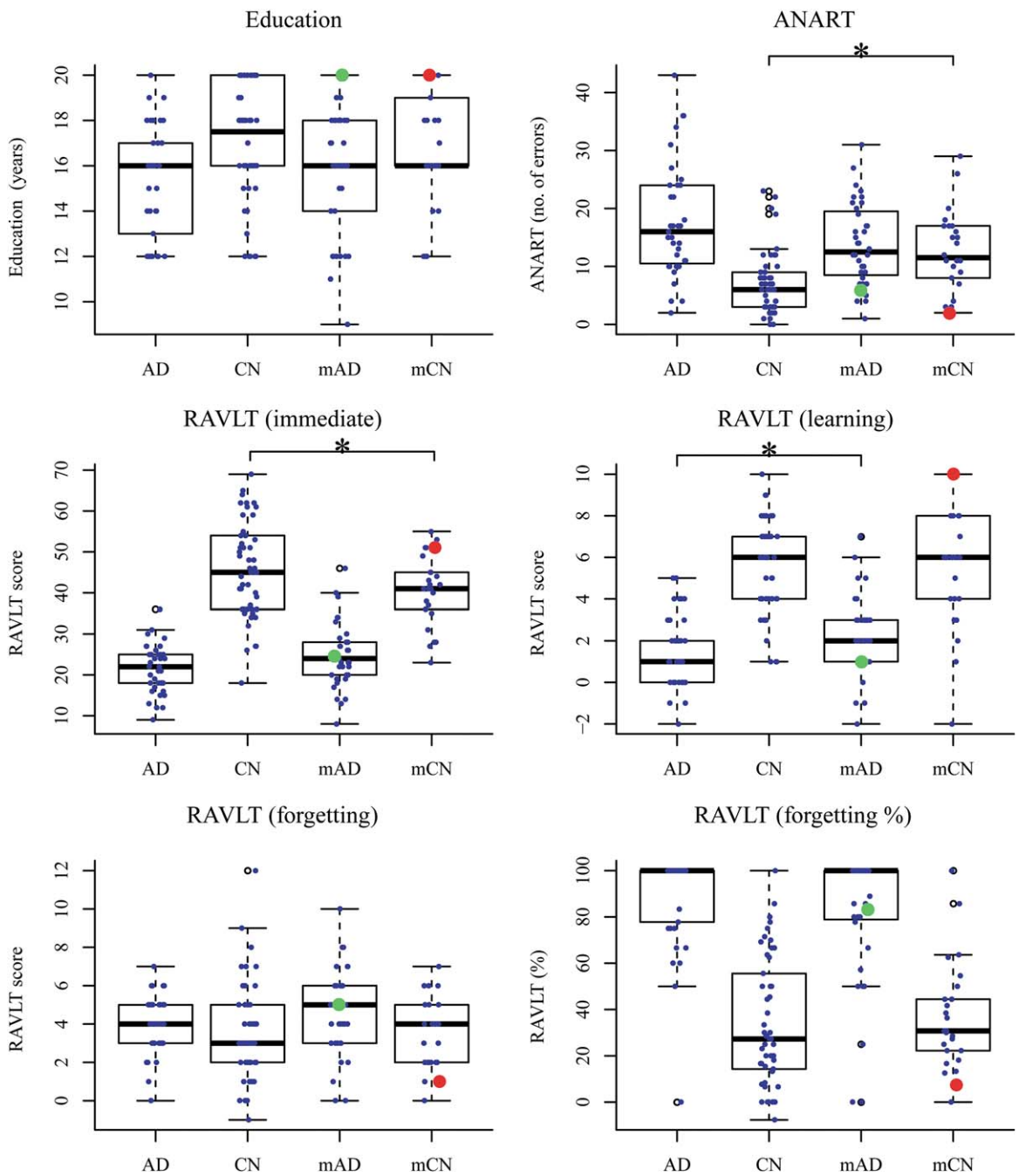

FIGURE 4 Plots showing the education level, ANART, and RVALT scores. Single misclassified control subject (ADNI subject Id: 127_S_5185) is represented as red circle and a single misclassified AD subject (ADNI subject Id: 052_S_4959) is represented as green circle. The symbol * indicates significantly different. AD = classified AD, mAD = misclassified AD, CN = classified control, mCN = misclassified control [Color figure can be viewed at wileyonlinelibrary.com]

At the group level, CSF tau and amyloid-beta were not significantly different between classified and misclassified subjects within each group type (in AD and in controls compared to misclassified within group). Analysis of WMH volume suggested that misclassified controls had a significantly greater WMH burden compared to correctly classified controls ($p = .011$, by two-sample t test) (Table 2). We therefore performed an additional classifier analysis incorporating white matter hyperintensity features with the initial classifiers. An accuracy of 91%, sensitivity of 0.98, and specificity 0.75 was obtained using the SVM classifier. Results suggest that inclusion of WMH features doesn't have a major impact on classification performance; however, there are several ways that WMH could provide additional classification information including use of spatial properties which differ in individuals with AD compared to controls. Thus, although these factors differ in the misclassified groups to some degree, the degree of heterogeneity prevents useful simple incorporation into the SVM classification procedure although future work will explore optimal parameters necessary for such incorporation.

TABLE 2 Two-sample *t* test analysis between classified and misclassified subjects within each group type

	Classified controls vs misclassified controls, *p* value	Classified AD vs misclassified AD, *p* value
Tau	.728	.990
Amyloid-beta	.564	.195
WMH	.011*	.298
MMSE	.045*	.272
Education	.777	.346
ANART	.001*	.056
RAVLT (immediate)	.048*	.054
RAVLT (learning)	.883	.042*
RAVLT (forgetting)	.969	.213
RAVLT (forgetting %)	.754	.253

Note. * Significantly different.

Analysis of MMSE score revealed significant difference between the correctly classified controls and misclassified controls ($p = .045$). The distribution of APOE4 (positive/negative for at least one e4 allele) and (18) F-AV45 uptake (standardized uptake values ratio (SUVR) > 1.11: florbetapir positive otherwise negative) were not different between the groups. We examined the influence of scanner types on the groups and found that distribution of Siemens and Philips scanners were not different between the classified and misclassified subjects. The variations attributable to individual scanner types may have less effect on the groups except for a greater representation of GE scanner type in the misclassified control group as shown in Figure 3. Although these data provide information about trends at the group level, within group variability across measures suggest heterogeneity supporting additional analyses of individual subjects from the sample.

ANART score, typically used as an estimate of premorbid verbal intelligence and serves as a proxy of cognitive reserve (Katzman et al., 1988; Lo and Jagust, 2013; McGurn et al., 2004; Schmand, Smit, Geerlings, & Lindeboom, 1997; Stern et al., 1994; Stern, 2012). ANART score was significantly different between the misclassified controls and correctly classified controls ($p = .001$). Analysis of RAVLT score revealed immediate recall scores were significantly different between the correctly classified controls and misclassified controls ($p = .048$) and also RAVLT learning scores were significantly different between the correctly classified AD and misclassified AD ($p = .042$). The RAVLT forgetting score, percentage of forgetting, and education level were not significantly different between classified and misclassified subjects. In summary, the misclassified controls had higher ANART and poorer RAVLT performance relative to correctly classified controls. These results suggest that some portion of the misclassified controls may in fact be in early stages of impairment, potentially masked due to higher premorbid function. Misclassified AD may be in the earlier stages of impairment relative to correctly classified AD. Although the thinning pattern is indicative of early AD pathology, it should be noted that misclassified AD did not differ from correctly classified AD with regard to biomarker levels or cognition. It is unclear why this is, however, it is possible that the cortical thickness measures are more sensitively quantifiable than the cognitive or CSF values (i.e., a range of thickness values related to variation in pathology are linked to similar cognitive and CSF values). Alternatively, it is possible that the differences in thickness are linked to subtle differences in white matter lesion volumes, or similarly, that the misclassified AD are generally healthier in various other domains that may influence cortical thickness whereas the typical AD patient is less healthy generally.

At the individual subject level, the misclassified control (misclassified control with the highest CSF tau levels across all misclassified controls) had MMSE of 30, APOE4 positive, (18) F-AV45 positive, 0.51% white matter hyperintesity, 132 pg/ml amyloid-beta, and 156 pg/ml tau. One sample *t* test presented in Table 3 showed significant difference between the single misclassified control subject and correctly classified controls across several parameters including increased CSF concentration of tau (selected for this but in the misclassified group), decreased concentration of CSF amyloid-beta, increased WMH, APOE4 and (18) F-AV45 uptake positivity, all considered to be typical characteristic factors related to presence or risk for AD. The misclassified control had an education of 20 years, ANART total score (no of errors) of 2, RAVLT immediate recall score (total of 5 trails) of 51, learning score (6th trial) of 10, forgetting score (30 min delay) of 1, and percentage of forgetting 7.14%. RAVLT score in this individual suggested subtle impairment, however, this individual did not convert to MCI or AD during the course of ADNI participation (from baseline visit to 24 months visit). These consistent results indicates the possibility of clinical misdiagnosis relative to pathologically positive AD in certain individuals in the control group, likely due substantially to "cognitive reserve" based in high education and premorbid intelligence. The individual misclassified AD had lower hyperintensity burden, and greater MMSE than correctly classified AD. The individual had high education and premorbid intelligence, which likely contributed to some preserved cognitive capacity. Taken together, it is possible that this individual is impaired for reasons other than AD pathology.

The misclassified AD subject had MMSE of 25, was APOE4 positive, (18) F-AV45 positive, 119 pg/ml amyloid-beta, and 42 pg/ml tau, but WMH of 0.47%. The misclassified AD had an education of 20 years, ANART score of 6, RAVLT immediate recall score of 26, learning score of 1, forgetting score of 5, and percentage of forgetting 83.33%. The misclassified AD remained fairly stable cognitively (memory loss prominent with poor insight) and functionally from baseline visit throughout their participation in ADNI (to 12 months visit). This example is somewhat inconsistent in having low tau but also low amyloid CSF values (tau levels would be expected to be high and amyloid low in AD). Findings suggest that given the high general prevalence of WMH in AD, patients with less WMH are likely to show altered patterns of cortical thinning compared to those who do not have WMH which may contribute to misclassification. Overall, the findings from the individual participant analyses demonstrate that individuals within the sample have clusters of demographic, biomarker, and neuropsychological profiles that are inconsistent with a match between the clinical

TABLE 3 One-sample *t* test analysis between correctly classified group and misclassified single subject within each group type

	Misclassified single control subject	Misclassified single AD subject	Correctly classified control group vs misclassified single control subject, *p* value	Correctly classified AD group vs misclassified single AD subject, *p* value
Tau (pg/ml)	156	42	<.001	<.001
Amyloid-beta (pg/ml)	132	119	<.001	<.001
WMH (%)	0.51	0.47	<.001	>.05
MMSE	30	25	<.001	<.001
Education	20	20	<.001	<.001
ANART	2	6	<.001	<.001
RAVLT (immediate)	51	26	<.001	<.001
RAVLT (learning)	10	1	<.001	>.05
RAVLT (forgetting)	1	5	<.001	<.001
RAVLT (forgetting %)	7.14	83.33	<.001	>.05

diagnosis and likely existence of AD pathology and also highlight factors that modulate the efficacy of the structural classification.

4 | DISCUSSION

There is strong interest in AD "diagnostics" using machine learning and structural MRI (Cho et al., 2012; Coupé et al., 2012; Davatzikos, Bhatt, Shaw, Batmanghelich, & Trojanowski, 2011; Falahati et al., 2014; Klöppel et al., 2015; Liu et al., 2012; Schouten et al., 2016; Westman et al., 2011a, 2011b; Westman, Muehlboeck, & Simmons, 2012; Zhang et al., 2011). These studies have demonstrated the obvious utility of structural imaging in this endeavor. Several studies have shown that hippocampal atrophy is an early indicator of AD (Jack et al., 1999; Killiany et al., 2002; Rana et al., 2017; Schröder and Pantel, 2016). The hippocampus has therefore been used in several prior classification studies (Chupin et al., 2009; Colliot et al., 2008; Frisoni et al., 1999) and is very effective to achieving classification to similar degrees as other structural measures. However, most prior studies aimed to enhance classification accuracy relative to clinical diagnosis through the optimal selection of classifiers for matching the clinical label. Little is currently known about factors that contribute to misclassification of AD purely from structural MRI which should mirror the primary *pathologies* of AD. The goal of this work was to determine factors that contribute to the mismatch between structural classification and clinical diagnosis through examination of a range of demographic, biological, and neuropsychological data. The overall findings demonstrate that in fact, although a clinical misclassification of ~10% was found, the structural classifier may actually have greater accuracy for a pathologic diagnosis (which would be most useful for clinical trials and therapeutics directed towards primary AD pathology), and also that secondary pathologies (such as WMH) influence thickness values in regions overlapping typical AD pathology and therefore the accuracy of classification.

Cortical thickness measurements based on MRI has been shown to index pathology in AD (Bakkour et al., 2009; Dickerson et al., 2009; Lerch et al., 2008; McDonald et al., 2009; Salat et al., 2011). In this study, as expected based on the use of cortical thickness measures in the SVM classifier, whole surface contrasts of cortical thickness maps between classified controls and misclassified controls revealed reduced cortical thickness in misclassified controls compared to correctly classified controls as shown in Figure 2. However, the regional patterns of reduced thickness were somewhat distinct from regions showing strong effects for AD including the regions used in training the classifier and seemed to be more global in nature, at least at the group level.

The need for integration of biomarker information into the diagnosis of individuals with pathologic AD has been discussed extensively (Jack et al., 2016; Mattsson et al., 2015; Palmqvist et al., 2015). These consensus papers suggest that amyloid-beta biomarkers, tau biomarkers, and biomarkers of neurodegeneration are necessary to identify early AD with high accuracy and would be useful to understand disease pathogenesis and expedite drug development. WMH are common type of brain tissue alteration in older adults, more prevalent in AD and are related to brain structural measures and therefore this tissue damage is an important factor related to diagnosis of AD. Several studies have investigated the association of WMH and cortical atrophy and generally found a higher degree of cortical atrophy among individuals with higher burden of WMH (Appelman et al., 2009; Capizzano et al., 2004; Godin et al., 2009; Raji et al., 2012). There is increasing evidence of WMH association with cognitive decline (Prins and Scheltens, 2015; Provenzano et al., 2013; Rieckmann et al., 2016). Recent studies have demonstrated interactions between WMH and cortical thickness and cognition (Jacobs, Clerx, Gronenschild, Aalten, & Verhey, 2014; Seo et al., 2012; Tuladhar et al., 2015). In this study, we found misclassified individuals differed from their correctly classified counterparts on several of these relevant measures. For example, misclassified controls, as a group, had a significantly greater WMH burden compared to correctly classified controls. The controls with more WMH are likely to show altered patterns of cortical thinning compared to those who do not have WMH which may contribute to misclassification. These findings suggest that WMH may be used as an additional biomarker for early and accurate diagnosis of AD. Ongoing studies are exploring the

utility of WMH in classification of AD, however, given that our goal here was to classify based on features considered to be linked to the primary pathology of AD, the inclusion of WMH to better match the clinical diagnosis would not necessarily help in achieving the goal of a pathologic classification.

Education and premorbid verbal intelligence typically serves as a proxy of cognitive reserve. Prior studies have shown association of the cognitive reserve markers with a lower risk of AD and memory decline (Buckner, 2004; Katzman et al., 1988; Lo and Jagust, 2013; Murray et al., 2011; Stern et al., 1994; Stern, 2012). Premorbid intelligence assessed by ANART modifies the relationship between biomarkers of pathology and cognition in AD with individuals with high cognitive reserve having greater biomarker abnormalities than those with low cognitive reserve (Vemuri et al., 2011). This study demonstrates misclassified controls had a significantly greater ANART, RAVLT immediate recall, and MMSE score compared to correctly classified controls. Thus, cognitive factors have a compensatory function that acts to preserve a clinical state despite reduced cortical thickness in some control individuals. It is therefore possible that the structural MRI measures provide an accurate pathologic, but not clinical diagnosis.

One limitation of the current work is that the classification procedure relied on clinical diagnostic information for determination of features selected. Although the large majority of participants likely have a match of their clinical diagnosis with brain pathology, it is possible that some inaccuracies in diagnosis shifted performance of the classifier to some degree. Future investigations should consider creation of a classifier based on "high confidence" individuals with diagnoses that are strongly supported by genetic and CSF biomarker data for optimal weighting towards classification based on AD pathologic processes. In the current investigation, a single misclassified control subject was selected based on having the greatest CSF tau levels in that group and a single misclassified AD subject was selected based on having the lowest CSF tau levels in that group for example comparisons. These single subjects are not presented as representative of the group, but simply as an example of the individual variability within the misclassified group. Although we highlight two individuals as an example, full subject data for all subjects classified and misclassified is provided in Figures 3 and 4. These plots show trends in the data at the group level. As noted, there are a range of values for correctly classified and misclassified individuals. This suggests that there is more than one explanation for misclassification as we tried to touch on in the current work. A secondary procedure such as hierarchical clustering could potentially subclassify the misclassified individuals and provide insight to the types of misclassification that occur in the sample as follow up investigation to the current work.

The major conclusions from current work are as follows: (a) WMH have an important influence on classification accuracy, with individuals with a clinical diagnosis of AD being more likely to be classified as control when WMH volume is low, and control individuals being more likely to be classified as AD when WMH volume is high. (b) At an individual level, subjects in the ADNI sample have biomarker evidence of AD pathology while remaining relatively cognitively resilient and having a clinical diagnosis of being nondemented. (c) Individuals clinically diagnosed with AD but misclassified based on cortical thickness patterns may have a pathologic diagnosis inconsistent with AD. These findings demonstrate the need for integration of biomarker based diagnostic criteria as has been described in recent consensus papers and recent classification work (Jack et al., 2016; Mattsson et al., 2015; Palmqvist et al., 2015). Here we note the critical contribution of white matter lesions to biomarker interpretation. A more detailed analysis of individual level factors across many subjects in the sample will be important follow up work to the data presented here. Thus, patterns of cortical thinning alone cannot be the only features to be used in clinical classification schemes. In the absence of additional biomarker data, some information regarding premorbid function and WMH burden must likely be included in these procedures. Additionally, future investigation should be explicit regarding goals of clinical versus pathologic diagnostic accuracy. Future work will examine the degree to which subgroups of individuals can be determined from the misclassified individuals based on clustering of AD and demographic related parameters to better understand features classes that contribute to classification accuracy.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health – NIH (grant number R01NR010827, using resources provided by NIH grants NS042861, NS058793); by the Center for Functional Neuroimaging Technologies (P41RR14075), a P41 Regional Resource supported by the Biomedical Technology Program of the National Center for Research Resources (NCRR), NIH; and by the NCRR Shared Instrumentation Grant Program and/or High-End Instrumentation Grant Program (grant numbers S10RR021110, S10RR023401, S10RR019307, S10RR019254, and S10RR023043). Dr. Mahanand received support from Raman Fellowship awarded by University Grants Commission, Government of India. The authors would like to thank Joost M. Riphagen, Emily R. Lindemer, and Douglas N. Greve for their useful suggestions.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen;Bristol-MyersSquibb Company; CereSpir, Inc.; Cogstate;Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health

Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

ORCID

Mahanand Belathur Suresh http://orcid.org/0000-0001-8189-6889

REFERENCES

Aguilar, C., Westman, E., Muehlboeck, J.-S., Mecocci, P., Vellas, B., Tsolaki, M., ... Wahlund, L.-O. (2013). Different multivariate techniques for automated classification of MRI data in Alzheimer's disease and mild cognitive impairment. *Psychiatry Research*, *212*, 89–98. https://doi.org/10.1016/j.pscychresns.2012.11.005

Appelman, A. P. A., Exalto, L. G., van der Graaf, Y., Biessels, G. J., Mali, W. P. T. M., & Geerlings, M. I. (2009). White matter lesions and brain atrophy: More than shared risk factors? A systematic review. *Cerebrovascular Diseases*, *28*, 227–242. https://doi.org/10.1159/000226774

Arnold, S. E., Hyman, B. T., Flory, J., Damasio, A. R., & Van Hoesen, G. W. (1991). The topographical and neuroanatomical distribution of neurofibrillary tangles and neuritic plaques in the cerebral cortex of patients with Alzheimer's disease. *Cerebral Cortex*, *1*, 103–116.

Bäckman, L., Jones, S., Berger, A.-K., Laukka, E. J., & Small, B. J. (2004). Multiple cognitive deficits during the transition to Alzheimer's disease. *Journal of Internal Medicine*, *256*, 195–204. https://doi.org/10.1111/j.1365-2796.2004.01386.x

Bakkour, A., Morris, J. C., & Dickerson, B. C. (2009). The cortical signature of prodromal AD Regional thinning predicts mild AD dementia. *Neurology*, *72*, 1048–1055.

Beach, T. G., Monsell, S. E., Phillips, L. E., & Kukull, W. (2012). Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005–2010. *Journal of Neuropathology & Experimental Neurology*, *71*, 266–273. https://doi.org/10.1097/NEN.0b013e31824b211b

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

Brun, A., & Gustafson, L. (1976). Distribution of cerebral degeneration in Alzheimer's disease. A clinico-pathological study. *Archiv fur Psychiatrie Und Nervenkrankheiten*, *223*, 15–33. (1970)

Buckner, R. L. (2004). Memory and executive function in aging and AD: Multiple factors that cause decline and reserve factors that compensate. *Neuron*, *44*, 195–208. https://doi.org/10.1016/j.neuron.2004.09.006

Burns, A., & Iliffe, S. (2009). Alzheimer's disease. *BMJ (Clinical Research Ed.)*, *338*, b158–b158. https://doi.org/10.1136/bmj.b158

Capizzano, A. A.,Ación, L., Bekinschtein, T., Furman, M., Gomila, H., Martínez, A., ... Starkstein, S. E. (2004). White matter hyperintensities are significantly associated with cortical atrophy in Alzheimer's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, *75*, 822–827.

Carlesimo, G. A., & Oscar-Berman, M. (1992). Memory deficits in Alzheimer's patients: A comprehensive review. *Neuropsychology Review*, *3*, 119–169.

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, *2*, 27.

Chetelat, G., & Baron, J.-C. (2003). Early diagnosis of Alzheimer's disease: Contribution of structural neuroimaging. *NeuroImage*, *18*, 525–541.

Cho, Y., Seong, J.-K., Jeong, Y., & Shin, S. Y. Alzheimer's Disease Neuroimaging Initiative (2012). Individual subject classification for Alzheimer's disease based on incremental learning using a spatial frequency representation of cortical thickness data. *NeuroImage*, *59*, 2217–2230., https://doi.org/10.1016/j.neuroimage.2011.09.085

Chupin, M., Gérardin, E., Cuingnet, R., Boutet, C., Lemieux, L., Lehéricy, S., ... Colliot, O. Alzheimer's Disease Neuroimaging Initiative, (2009). Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI. *Hippocampus*, *19*, 579–587. https://doi.org/10.1002/hipo.20626

Colliot, O., Chételat, G., Chupin, M., Desgranges, B., Magnin, B., Benali, H., ... Lehéricy, S. (2008). Discrimination between Alzheimer disease, mild cognitive impairment, and normal aging by using automated segmentation of the hippocampus. *Radiology*, *248*, 194–201. https://doi.org/10.1148/radiol.2481070876

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*, 273–297.

Coupé, P., Eskildsen, S. F., Manjón, J. V., Fonov, V. S., Collins, D. L., & Alzheimer's Disease, N. I. (2012). Simultaneous segmentation and grading of anatomical structures for patient's classification: Application to Alzheimer's disease. *NeuroImage*, *59*, 3736–3747. https://doi.org/10.1016/j.neuroimage.2011.10.080

Coutu, J.-P., Goldblatt, A., Rosas, H. D., & Salat, D. H. Alzheimer's Disease Neuroimaging Initiative (ADNI), (2016). White matter changes are associated with ventricular expansion in aging, mild cognitive impairment, and Alzheimer's disease. *Journal of Alzheimer's Disease*, *49*, 329–342. https://doi.org/10.3233/JAD-150306

Coutu, J.-P., Lindemer, E. R., Konukoglu, E., & Salat, D. H. Alzheimer's Disease Neuroimaging Initiative (ADNI), (2017). Two distinct classes of degenerative change are independently linked to clinical progression in mild cognitive impairment. *Neurobiology of Aging*, *54*, 1–9. https://doi.org/10.1016/j.neurobiolaging.2017.02.005

Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehéricy, S., Habert, M.-O., ... Initiative, A. D. N. (2011). Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database. *NeuroImage*, *56*, 766–781.

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

Dale, A. M., & Sereno, M. I. (1993). Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: A linear approach. *Journal of Cognitive Neuroscience*, *5*, 162–176.

Davatzikos, C., Bhatt, P., Shaw, L. M., Batmanghelich, K. N., & Trojanowski, J. Q. (2011). Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. *Neurobiology of Aging*, *32*, 2322.e19–2327. https://doi.org/10.1016/j.neurobiolaging.2010.05.023

de Leeuw, F.-E. (2001). Prevalence of cerebral white matter lesions in elderly people: A population based magnetic resonance imaging study. The Rotterdam Scan Study. *Journal of Neurology, Neurosurgery & Psychiatry*, *70*, 9–14. https://doi.org/10.1136/jnnp.70.1.9

de Vos, F., Schouten, T. M., Hafkemeijer, A., Dopper, E. G. P., van Swieten, J. C., de Rooij, M., ... Rombouts, S. A. R. B. (2016). Combining multiple anatomical MRI measures improves Alzheimer's disease

classification. *Human Brain Mapping*, *37*, 1920–1929. https://doi.org/10.1002/hbm.23147

Debette, S., & Markus, H. S. (2010). The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: Systematic review and meta-analysis. *BMJ*, *341*, c3666–c3666. https://doi.org/10.1136/bmj.c3666

DeCarli, C., Miller, B. L., Swan, G. E., Reed, T., Wolf, P. A., Garner, J., ... Carmelli, D. (1999). Predictors of brain morphology for the men of the NHLBI twin study. *Stroke*, *30*, 529–536.

DeCarli, C., Murphy, D. G., Teichberg, D., Campbell, G., & Sobering, G. S. (1996). Local histogram correction of MRI spatially dependent image pixel intensity nonuniformity. *Journal of Magnetic Resonance Imaging*, *6*, 519–528.

DeCarli, C., Pauline, M., & Evan, F. (2013). Four tissue segmentation in ADNI II.

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., ... Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, *31*, 968– 980. https://doi.org/DOI: 10.1016/j.neuroimage.2006.01.021

Dickerson, B. C., Bakkour, A., Salat, D. H., Feczko, E., Pacheco, J., Greve, D. N., ... Rosas, H. D. (2009). The cortical signature of Alzheimer's disease: Regionally specific cortical thinning relates to symptom severity in very mild to mild AD dementia and is detectable in asymptomatic amyloid-positive individuals. *Cerebral Cortex*, *19*, 497–510.

Dubois, B., Feldman, H. H., Jacova, C., Hampel, H., Molinuevo, J. L., Blennow, K., ... Cummings, J. L. (2014). Advancing research diagnostic criteria for Alzheimer's disease: The IWG-2 criteria. *Lancet Neurology*, *13*, 614–629. https://doi.org/10.1016/S1474-4422(14)70090-0

Dukart, J., Schroeter, M. L., & Mueller, K. Alzheimer's Disease Neuroimaging Initiative, (2011). Age correction in dementia–matching to a healthy brain. *PLoS ONE*, *6*, e22193. https://doi.org/10.1371/journal.pone.0022193

Eskildsen, S. F., Coupé, P., García-Lorenzo, D., Fonov, V., Pruessner, J. C., Collins, D. L., & Initiative, A. D. N. (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., ... Westman, E. for the AddNeuroMed consortium and the Alzheimer's Disease Neuroimaging Initiative, (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 Topography*, *29*, 296–307. https://doi.org/10.1007/s10548-015-0455-1

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*, 685–708. https://doi.org/10.3233/JAD-131928

Fischl, B., & Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences of the United States of America*, *97*, 11050–11055. https://doi.org/10.1073/pnas.200033797

Fischl, B., Liu, A., & Dale, A. M. (2001). Automated manifold surgery: Constructing geometrically accurate and topologically correct models of the human cerebral cortex. *IEEE Transactions on Medical Imaging*, *20*, 70–80.

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

Fischl, B., Salat, D. H., van der Kouwe, A. J., Makris, N., Segonne, F., Quinn, B. T., & Dale, A. M. (2004a). Sequence-independent segmentation of magnetic resonance images. *NeuroImage*, *23*(Suppl 1), S69–S84. https://doi.org/10.1016/j.neuroimage.2004.07.016

Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *NeuroImage*, *9*, 195–207. https://doi.org/10.1006/nimg.1998.0396

Fischl, B., Sereno, M. I., Tootell, R. B. H., & Dale, A. M. (1999). High-resolution inter-subject averaging and a coordinate system for the cortical surface. *Human Brain Mapping*, *8*, 272–284.

Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Segonne, F., Salat, D. H., ... Dale, A. M. (2004b). Automatically parcellating the human cerebral cortex. *Cerebral Cortex*, *14*, 11–22.

Fotenos, A. F., Snyder, A. Z., Girton, L. E., Morris, J. C., & Buckner, R. L. (2005). Normative estimates of cross-sectional and longitudinal brain volume decline in aging and AD. *Neurology*, *64*, 1032–1039. https://doi.org/10.1212/01.WNL.0000154530.72969.11

Frisoni, G. B., Fox, N. C., Jack, C. R., Scheltens, P., & Thompson, P. M. (2010). The clinical use of structural MRI in Alzheimer disease. *Nature Reviews. Neurology*, *6*, 67–77. https://doi.org/10.1038/nrneurol.2009.215

Frisoni, G. B., Laakso, M. P., Beltramello, A., Geroldi, C., Bianchetti, A., Soininen, H., & Trabucchi, M. (1999). Hippocampal and entorhinal cortex atrophy in frontotemporal dementia and Alzheimer's disease. *Neurology*, *52*, 91–100.

Godin, O., Maillard, P., Crivello, F., Alpérovitch, A., Mazoyer, B., Tzourio, C., & Dufouil, C. (2009). Association of white-matter lesions with brain atrophy markers: The three-city Dijon MRI study. *Cerebrovascular Diseases*, *28*, 177–184. https://doi.org/10.1159/000226117

Hagler, D. J., Jr., Saygin, A. P., & Sereno, M. I. (2006). Smoothing and cluster thresholding for cortical surface-based group analysis of fMRI data. *NeuroImage*, *33*, 1093–1103. https://doi.org/10.1016/j.neuroimage.2006.07.036

Han, X., Jovicich, J., Salat, D., van der Kouwe, A., Quinn, B., Czanner, S., ... Fischl, B. (2006). Reliability of MRI-derived measurements of human cerebral cortical thickness: The effects of field strength, scanner upgrade and manufacturer. *NeuroImage*, *32*, 180–194. https://doi.org/10.1016/j.neuroimage.2006.02.051

Hardy, J. (2006). Alzheimer's disease: The amyloid cascade hypothesis: An update and reappraisal. *Journal of Alzheimer's Disease*, *9*, 151–153.

Holtzman, D. M., Morris, J. C., & Goate, A. M. (2011). Alzheimer's disease: The challenge of the second century. *Science Translational Medicine*, *3*, 77sr1. https://doi.org/10.1126/scitranslmed.3002369

Hopkins, R. O., Beck, C. J., Burnett, D. L., Weaver, L. K., Victoroff, J., & Bigler, E. D. (2006). Prevalence of white matter hyperintensities in a young healthy population. *Journal of Neuroimaging*, *16*, 243–251. https://doi.org/10.1111/j.1552-6569.2006.00047.x

Hwang, J., Kim, C. M., Jeon, S., Lee, J. M., Hong, Y. J., Roh, J. H., ... Na, D. L. (2016). Prediction of Alzheimer's disease pathophysiology based on cortical thickness patterns. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *2*, 58–67. https://doi.org/10.1016/j.dadm.2015.11.008

Hyman, B. T., Phelps, C. H., Beach, T. G., Bigio, E. H., Cairns, N. J., Carrillo, M. C., ... Montine, T. J. (2012). National Institute on Aging–Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease. *Alzheimer's & Dementia*, *8*, 1–13. https://doi.org/10.1016/j.jalz.2011.10.007

Jack, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Feldman, H. H., Frisoni, G. B., ... Dubois, B. (2016). A/T/N: An unbiased descriptive

classification scheme for Alzheimer disease biomarkers. *Neurology*, *87*, 539–547. https://doi.org/10.1212/WNL.0000000000002923

Jack, C. R., Petersen, R. C., O'brien, P. C., & Tangalos, E. G. (1992). MR-based hippocampal volumetry in the diagnosis of Alzheimer's disease. *Neurology*, *42*, 183–188.

Jack, C. R., Petersen, R. C., Xu, Y. C., O'brien, P. C., Smith, G. E., Ivnik, R. J., ... Kokmen, E. (1999). Prediction of AD with MRI-based hippocampal volume in mild cognitive impairment. *Neurology*, *52*, 1397–1403.

Jack, C. R., Petersen, R. C., Xu, Y. C., Waring, S. C., O'brien, P. C., Tangalos, E. G., ... Kokmen, E. (1997). Medial temporal atrophy on MRI in normal aging and very mild Alzheimer's disease. *Neurology*, *49*, 786–794.

Jacobs, H. I. L., Clerx, L., Gronenschild, E. H. B. M., Aalten, P., & Verhey, F. R. J. (2014). White matter hyperintensities are positively associated with cortical thickness in Alzheimer's disease. *Journal of Alzheimer's Disease*, *39*, 409–422. https://doi.org/10.3233/JAD-131232

Joachim, C. L., Morris, J. H., & Selkoe, D. J. (1988). Clinically diagnosed Alzheimer's disease: Autopsy results in 150 cases. *Annals of Neurology*, *24*, 50–56. https://doi.org/10.1002/ana.410240110

Jovicich, J., Czanner, S., Greve, D., Haley, E., van der Kouwe, A., Gollub, R., ... Dale, A. (2006). Reliability in multi-site structural MRI studies: Effects of gradient non-linearity correction on phantom and human data. *NeuroImage*, *30*, 436–443. https://doi.org/10.1016/j.neuroimage.2005.09.046

Katzman, R., Terry, R., DeTeresa, R., Brown, T., Davies, P., Fuld, P., ... Peck, A. (1988). Clinical, pathological, and neurochemical changes in dementia: A subgroup with preserved mental status and numerous neocortical plaques. *Annals of Neurology*, *23*, 138–144. https://doi.org/10.1002/ana.410230206

Killiany, R. J., Gomez-Isla, T., Moss, M., Kikinis, R., Sandor, T., Jolesz, F., ... Albert, M. S. (2000). Use of structural magnetic resonance imaging to predict who will get Alzheimer's disease. *Annals of Neurology*, *47*, 430–439.

Killiany, R. J., Hyman, B. T., Gomez-Isla, T., Moss, M. B., Kikinis, R., Jolesz, F., ... Albert, M. S. (2002). MRI measures of entorhinal cortex vs hippocampus in preclinical AD. *Neurology*, *58*, 1188–1196.

Klöppel, S., Peter, J., Ludl, A., Pilatus, A., Maier, S., Mader, I., ... Abdulkadir, A. Alzheimer's Disease Neuroimaging Initiative, (2015). Applying automated MR-based diagnostic methods to the memory clinic: A prospective study. *Journal of Alzheimer's Disease*, *47*, 939–954. https://doi.org/10.3233/JAD-150334

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

Kuperberg, G. R., Broome, M., McGuire, P. K., David, A. S., Eddy, M., Ozawa, F., ... Fischl, B. (2003). Regionally localized thinning of the cerebral cortex in Schizophrenia. *Archives of General Psychiatry*, *60*, 878–888.

Landau, S. M., Harvey, D., Madison, C. M., Reiman, E. M., Foster, N. L., Aisen, P. S., ... Jagust, W. J. Alzheimer's Disease Neuroimaging Initiative, (2010). Comparing predictors of conversion and decline in mild cognitive impairment. *Neurology*, *75*, 230–238. https://doi.org/10.1212/WNL.0b013e3181e8e8b8

Lerch, J. P., Pruessner, J., Zijdenbos, A. P., Collins, D. L., Teipel, S. J., Hampel, H., & Evans, A. C. (2008). Automated cortical thickness measurements from MRI can accurately separate Alzheimer's patients from normal elderly controls. *Neurobiology of Aging*, *29*, 23–30.

Lim, K. O., Zipursky, R. B., Murphy, G. M., & Pfefferbaum, A. (1990). In vivo quantification of the limbic system using MRI: Effects of normal aging. *Psychiatry Research*, *35*, 15–26.

Liu, M., Zhang, D., & Shen, D. Alzheimer's Disease Neuroimaging Initiative, (2012). Ensemble sparse classification of Alzheimer's disease. *NeuroImage*, *60*, 1106–1116. https://doi.org/10.1016/j.neuroimage.2012.01.055

Lo, R. Y., & Jagust, W. J. (2013). Effect of cognitive reserve markers on Alzheimer pathologic progression. *Alzheimer Disease & Associated Disorders*, *27*, 343–350. https://doi.org/10.1097/WAD.0b013e3182900b2b

Lopez, O. L., Becker, J. T., Klunk, W., Saxton, J., Hamilton, R. L., Kaufer, D. I., ... DeKosky, S. T. (2000). Research evaluation and diagnosis of probable Alzheimer's disease over the last two decades: I. *Neurology*, *55*, 1854–1862.

Magnin, B., Mesrob, L., Kinkingnéhun, S., Pélégrini-Issac, M., Colliot, O., Sarazin, M., ... Benali, H. (2009). Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. *Neuroradiology*, *51*, 73–83. https://doi.org/10.1007/s00234-008-0463-x

Mattsson, N., Carrillo, M. C., Dean, R. A., Devous, M. D., Nikolcheva, T., Pesini, P., ... Liu, E. (2015). Revolutionizing Alzheimer's disease and clinical trials through biomarkers. *Alzheimers & Dementia (Amsterdam)*, *1*, 412–419. https://doi.org/10.1016/j.dadm.2015.09.001

McDonald, C., McEvoy, L., Gharapetian, L., Fennema-Notestine, C., Hagler, D., & Holland, D. (2009). Regional rates of neocortical atrophy from normal aging to early Alzheimer disease. *Neurology*, *73*, 457–465.

McGurn, B., Starr, J. M., Topfer, J. A., Pattie, A., Whiteman, M. C., Lemmon, H. A., ... Deary, I. J. (2004). Pronunciation of irregular words is preserved in dementia, validating premorbid IQ estimation. *Neurology*, *62*, 1184–1186.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, *34*, 939–939. https://doi.org/10.1212/WNL.34.7.939

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Kawas, C. H., ... Phelps, C. H. (2011). The diagnosis of dementia 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*, 263–269. https://doi.org/10.1016/j.jalz.2011.03.005

Montine, T. J., Phelps, C. H., Beach, T. G., Bigio, E. H., Cairns, N. J., Dickson, D. W., ... Hyman, B. T. (2012). National Institute on Aging–Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease: A practical approach. *Acta Neuropathologica*, *123*, 1–11. https://doi.org/10.1007/s00401-011-0910-3

Mortamais, M., Reynes, C., Brickman, A. M., Provenzano, F. A., Muraskin, J., Portet, F., ... Artero, S. (2013). Spatial distribution of cerebral white matter lesions predicts progression to mild cognitive impairment and dementia. *PLoS ONE*, *8*, e56972. https://doi.org/10.1371/journal.pone.0056972

Murray, A. D., Staff, R. T., McNeil, C. J., Salarirad, S., Ahearn, T. S., Mustafa, N., & Whalley, L. J. (2011). The balance between cognitive reserve and brain imaging biomarkers of cerebrovascular and Alzheimer's diseases. *Brain*, *134*, 3687–3696. https://doi.org/10.1093/brain/awr259

Murray, A. D., Staff, R. T., Shenkin, S. D., Deary, I. J., Starr, J. M., & Whalley, L. J. (2005). Brain white matter hyperintensities: Relative importance of vascular risk factors in nondemented elderly people. *Radiology*, *237*, 251–257. https://doi.org/10.1148/radiol.2371041496

Palmqvist, S., Zetterberg, H., Mattsson, N., Johansson, P., Alzheimer's Disease Neuroimaging Initiative, Minthon, L., ... Hansson, O. Swedish BioFINDER Study Group, (2015). Detailed comparison of amyloid PET and CSF biomarkers for identifying early Alzheimer disease. *Neurology*, *85*, 1240–1249. https://doi.org/10.1212/WNL.0000000000001991

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 Journal of Radiology*, *17*, 827. https://doi.org/10.3348/kjr.2016.17.6.827

Petersen, R. C., Doody, R., Kurz, A., Mohs, R. C., Morris, J. C., Rabins, P. V., ... Winblad, B. (2001). Current concepts in mild cognitive impairment. *Archives of Neurology*, *58*, 1985–1992.

Prins, N. D., & Scheltens, P. (2015). White matter hyperintensities, cognitive impairment and dementia: An update. *Nature Reviews. Neurology*, *11*, 157–165. https://doi.org/10.1038/nrneurol.2015.10

Provenzano, F. A., Muraskin, J., Tosto, G., Narkhede, A., Wasserman, B. T., Griffith, E. Y., ... Brickman, A. M. (2013). White matter hyperintensities and cerebral amyloidosis: Necessary and sufficient for clinical expression of Alzheimer disease? *JAMA Neurology*, *70*, 455–461.

Querfurth, H. W., & LaFerla, F. M. (2010). Alzheimer's disease. *The New England Journal of Medicine*, *362*, 329–344. https://doi.org/10.1056/NEJMra0909142

Raamana, P. R., Weiner, M. W., Wang, L., Beg, M. F., & Alzheimer's Disease Neuroimaging Initiative. (2015). Thickness network features for prognostic applications in dementia. *Neurobiology of Aging*, *36*, S91–S102.

Raji, C. A., Lopez, O. L., Kuller, L. H., Carmichael, O. T., Longstreth, W. T., Gach, H. M., ... Becker, J. T. (2012). White matter lesions and brain gray matter volume in cognitively normal elders. *Neurobiology of Aging*, *33*, 834.e7–816. https://doi.org/10.1016/j.neurobiolaging.2011.08.010

Rana, A. K., Sandu, A.-L., Robertson, K. L., McNeil, C. J., Whalley, L. J., Staff, R. T., & Murray, A. D. (2017). A comparison of measurement methods of hippocampal atrophy rate for predicting Alzheimer's dementia in the Aberdeen Birth Cohort of 1936. *Alzheimers & Dementia (Amsterdam)*, *6*, 31–39. https://doi.org/10.1016/j.dadm.2016.11.007

Reuter, M., Rosas, H. D., & Fischl, B. (2010). Highly accurate inverse consistent registration: A robust approach. *NeuroImage*, *53*, 1181–1196. https://doi.org/10.1016/j.neuroimage.2010.07.020

Reuter, M., Schmansky, N. J., Rosas, H. D., & Fischl, B. (2012). Within-subject template estimation for unbiased longitudinal image analysis. *NeuroImage*, *61*, 1402–1418. https://doi.org/10.1016/j.neuroimage.2012.02.084

Rieckmann, A., Van Dijk, K. R. A., Sperling, R. A., Johnson, K. A., Buckner, R. L., & Hedden, T. (2016). Accelerated decline in white matter integrity in clinically normal individuals at risk for Alzheimer's disease. *Neurobiology of Aging*, *42*, 177–188. https://doi.org/10.1016/j.neurobiolaging.2016.03.016

Rosas, H. D., Liu, A. K., Hersch, S., Glessner, M., Ferrante, R. J., Salat, D. H., ... Fischl, B. (2002). Regional and progressive thinning of the cortical ribbon in Huntington's disease. *Neurology*, *58*, 695–701.

Salat, D. H., Buckner, R. L., Snyder, A. Z., Greve, D. N., Desikan, R. S., Busa, E., ... Fischl, B. (2004). Thinning of the cerebral cortex in aging. *Cerebral Cortex (New York, N.Y.: 1991)*, *14*, 721–730. https://doi.org/10.1093/cercor/bhh032

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*, 1795–1802. https://doi.org/10.1016/j.neuroimage.2010.10.034

Salat, D. H., Kaye, J. A., & Janowsky, J. S. (2001). Selective preservation and degeneration within the prefrontal cortex in aging and Alzheimer disease. *Archives of Neurology*, *58*, 1403–1408.

Salat, D. H., Kaye, J. A., & Janowsky, J. S. (1999). Prefrontal gray and white matter volumes in healthy aging and Alzheimer disease. *Archives of Neurology*, *56*, 338–344.

Salvatore, C., Battista, P., & Castiglioni, I. (2016). Frontiers for the early diagnosis of AD by means of MRI brain imaging and support vector machines. *Current Alzheimer Research*, *13*, 509–533.

Scheltens, P., Fox, N., Barkhof, F., & De Carli, C. (2002). Structural magnetic resonance imaging in the practical assessment of dementia: Beyond exclusion. *Lancet Neurology*, *1*, 13–21.

Scheltens, P., Leys, D., Barkhof, F., Huglo, D., Weinstein, H. C., Vermersch, P., ... Valk, J. (1992). Atrophy of medial temporal lobes on MRI in "probable" Alzheimer's disease and normal ageing: Diagnostic value and neuropsychological correlates. *Journal of Neurology, Neurosurgery & Psychiatry*, *55*, 967–972.

Schmand, B., Smit, J. H., Geerlings, M. I., & Lindeboom, J. (1997). The effects of intelligence and education on the development of dementia. A test of the brain reserve hypothesis. *Psychological Medicine*, *27*, 1337–1344.

Schmitter, D., Roche, A., Maréchal, B., Ribes, D., Abdulkadir, A., Bach-Cuadra, M., ... Krueger, G. Alzheimer's Disease Neuroimaging Initiative, (2015). An evaluation of volume-based morphometry for prediction of mild cognitive impairment and Alzheimer's disease. *NeuroImage. Clinical*, *7*, 7–17. https://doi.org/10.1016/j.nicl.2014.11.001

Schölkopf, B., & Smola, A. J. (2002). *Learning with kernels: Support vector machines, regularization, optimization, and beyond*. MIT Press.

Schouten, T. M., Koini, M., de Vos, F., Seiler, S., van der Grond, J., Lechner, A., ... Rombouts, S. A. R. B. (2016). Combining anatomical, diffusion, and resting state functional magnetic resonance imaging for individual classification of mild and moderate Alzheimer's disease. *NeuroImage. Clinical*, *11*, 46–51. https://doi.org/10.1016/j.nicl.2016.01.002

Schröder, J., & Pantel, J. (2016). Neuroimaging of hippocampal atrophy in early recognition of Alzheimer's disease–a critical appraisal after two decades of research. *Psychiatry Research*, *247*, 71–78. https://doi.org/10.1016/j.pscychresns.2015.08.014

Segonne, F., Dale, A. M., Busa, E., Glessner, M., Salat, D., Hahn, H. K., & Fischl, B. (2004). A hybrid approach to the skull stripping problem in MRI. *NeuroImage*, *22*, 1060–1075. https://doi.org/DOI: 10.1016/j.neuroimage.2004.03.032

Segonne, F., Pacheco, J., & Fischl, B. (2007). Geometrically accurate topology-correction of cortical surfaces using nonseparating loops. *IEEE Transactions on Medical Imaging*, *26*, 518–529.

Selkoe, D. J., & Hardy, J. (2016). The amyloid hypothesis of Alzheimer's disease at 25 years. *EMBO Molecular Medicine*, *8*, 595–608. https://doi.org/10.15252/emmm.201606210

Seo, S. W., Lee, J.-M., Im, K., Park, J.-S., Kim, S.-H., Kim, S. T., ... Na, D. L. (2012). Cortical thinning related to periventricular and deep white matter hyperintensities. *Neurobiology of Aging*, *33*, 1156–1167. https://doi.org/10.1016/j.neurobiolaging.2010.12.003

Shear, P. K., Sullivan, E. V., Mathalon, D. H., Lim, K. O., Davis, L. F., Yesavage, J. A., ... Pfefferbaum, A. (1995). Longitudinal volumetric computed tomographic analysis of regional brain changes in normal aging and Alzheimer's disease. *Archives of Neurology*, *52*, 392–402.

Sled, J. G., Zijdenbos, A. P., & Evans, A. C. (1998). A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Transactions on Medical Imaging*, *17*, 87–97.

Stern, Y. (2012). Cognitive reserve in ageing and Alzheimer's disease. *The Lancet. Neurology*, *11*, 1006–1012. https://doi.org/10.1016/S1474-4422(12)70191-6

Stern, Y., Gurland, B., Tatemichi, T. K., Tang, M. X., Wilder, D., & Mayeux, R. (1994). Influence of education and occupation on the incidence of Alzheimer's disease. *JAMA*, *271*, 1004–1010.

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. *The Medical Clinics of North America*, *97*, 399–424. https://doi.org/10.1016/j.mcna.2012.12.013

Thompson, P. M., Moussai, J., Zohoori, S., Goldkorn, A., Khan, A. A., Mega, M. S., . . . Toga, A. W. (1998). Cortical variability and asymmetry in normal aging and Alzheimer's disease. *Cerebral Cortex (New York, N.Y.: 1991)*, *8*, 492–509.

Tuladhar, A. M., Reid, A. T., Shumskaya, E., de Laat, K. F., van Norden, A. G. W., van Dijk, E. J., . . . de Leeuw, F.-E. (2015). Relationship between white matter hyperintensities, cortical thickness, and cognition. *Stroke*, *46*, 425–432. https://doi.org/10.1161/STROKEAHA.114.007146

Vemuri, P., Gunter, J. L., Senjem, M. L., Whitwell, J. L., Kantarci, K., Knopman, D. S., . . . Jack, C. R. (2008). Alzheimer's disease diagnosis in individual subjects using structural MR images: Validation studies. *NeuroImage*, *39*, 1186–1197.

Vemuri, P., Weigand, S. D., Przybelski, S. A., Knopman, D. S., Smith, G. E., Trojanowski, J. Q., . . . Jack, C. R. Alzheimer's Disease Neuroimaging Initiative, (2011). Cognitive reserve and Alzheimer's disease biomarkers are independent determinants of cognition. *Brain*, *134*, 1479–1492. https://doi.org/10.1093/brain/awr049

Westman, E., Cavallin, L., Muehlboeck, J.-S., Zhang, Y., Mecocci, P., Vellas, B., . . . Wahlund, L.-O. AddNeuroMed consortium, (2011a). Sensitivity and specificity of medial temporal lobe visual ratings and multivariate regional MRI classification in Alzheimer's disease. *PLoS ONE*, *6*, e22506. https://doi.org/10.1371/journal.pone.0022506

Westman, E., Muehlboeck, J.-S., & Simmons, A. (2012). Combining MRI and CSF measures for classification of Alzheimer's disease and prediction of mild cognitive impairment conversion. *NeuroImage*, *62*, 229–238. https://doi.org/10.1016/j.neuroimage.2012.04.056

Westman, E., Simmons, A., Muehlboeck, J.-S., Mecocci, P., Vellas, B., Tsolaki, M., . . . Wahlund, L.-O. AddNeuroMed consortium, Alzheimer's Disease Neuroimaging Initiative (2011b). AddNeuroMed and ADNI: Similar patterns of Alzheimer's atrophy and automated MRI classification accuracy in Europe and North America. *NeuroImage*, *58*, 818–828. https://doi.org/10.1016/j.neuroimage.2011.06.065

Wolz, R., Julkunen, V., Koikkalainen, J., Niskanen, E., Zhang, D. P., Rueckert, D., & Alzheimer's Disease Neuroimaging Initiative. (2011). Multimethod analysis of MRI images in early diagnostics of Alzheimer's disease. *PLoS ONE*, *6*, e25446.

Ylikoski, A., Erkinjuntti, T., Raininko, R., Sarna, S., Sulkava, R., & Tilvis, R. (1995). White matter hyperintensities on MRI in the neurologically nondiseased elderly. Analysis of cohorts of consecutive subjects aged 55 to 85 years living at home. *Stroke*, *26*, 1171–1177.

Zhang, D., Wang, Y., Zhou, L., Yuan, H., & Shen, D. Alzheimer's Disease Neuroimaging Initiative (2011). Multimodal classification of Alzheimer's disease and mild cognitive impairment. *NeuroImage*, *55*, 856–867. https://doi.org/10.1016/j.neuroimage.2011.01.008

How to cite this article: Belathur Suresh M, Fischl B, Salat DH. Factors influencing accuracy of cortical thickness in the diagnosis of Alzheimer's disease. *Hum Brain Mapp*. 2018;39:1500–1515. https://doi.org/10.1002/hbm.23922