Journal of Alzheimer's Disease xx (2021) x–xx
DOI 10.3233/JAD-200821
IOS Press

Improved Prediction of Imminent Progression to Clinically Significant Memory Decline Using Surface Multivariate Morphometry Statistics and Sparse Coding

Cynthia M. Stonnington[a,*], Jianfeng Wu[b], Jie Zhang[b], Jie Shi[b], Robert J. Bauer III[c], Vivek Devadas[c], Yi Su[c], Dona E.C. Locke[a], Eric M. Reiman[c], Richard J. Caselli[d], Kewei Chen[c] and Yalin Wang[b] for the Alzheimer's Disease Neuroimaging Initiative[1]
[a]*Department of Psychiatry and Psychology, Mayo Clinic, Scottsdale, AZ, USA*
[b]*School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe, AZ, USA*
[c]*Banner Alzheimer's Institute, Phoenix, AZ, USA*
[b]*Department of Neurology, Mayo Clinic, Scottsdale, AZ, USA*

Accepted 16 February 2021
Pre-press 13 March 2021

Abstract.
Background: Besides their other roles, brain imaging and other biomarkers of Alzheimer's disease (AD) have the potential to inform a cognitively unimpaired (CU) person's likelihood of progression to mild cognitive impairment (MCI) and benefit subject selection when evaluating promising prevention therapies. We previously described that among baseline FDG-PET and MRI measures known to be preferentially affected in the preclinical and clinical stages of AD, hippocampal volume was the best predictor of incident MCI within 2 years (79% sensitivity/78% specificity), using standard automated MRI volumetric algorithmic programs, binary logistic regression, and leave-one-out procedures.
Objective: To improve the same prediction by using different hippocampal features and machine learning methods, cross-validated via two independent and prospective cohorts (Arizona and ADNI).
Methods: Patch-based sparse coding algorithms were applied to hippocampal surface features of baseline TI-MRIs from 78 CU adults who subsequently progressed to amnestic MCI in approximately 2 years ("progressors") and 80 matched adults who remained CU for at least 4 years ("nonprogressors"). Nonprogressors and progressors were matched for age, sex, education, and apolipoprotein E4 allele dose. We did not include amyloid or tau biomarkers in defining MCI.
Results: We achieved 92% prediction accuracy in the Arizona cohort, 92% prediction accuracy in the ADNI cohort, and 90% prediction accuracy when combining the two demographically distinct cohorts, as compared to 79% (Arizona) and 72% (ADNI) prediction accuracy using hippocampal volume.
Conclusion: Surface multivariate morphometry and sparse coding, applied to individual MRIs, may accurately predict imminent progression to MCI even in the absence of other AD biomarkers.

Keywords: Alzheimer's disease, magnetic resonance imaging, mild cognitive impairment, prediction, prognosis

[1]Some of the data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://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

*Correspondence to: Cynthia M. Stonnington, M.D., Department of Psychiatry and Psychology, Mayo Clinic, 13400 E Shea Blvd., Scottsdale, AZ 85259, USA. Tel.: +1 480 301 4853; Fax: +1 480 301 6258; E-mail: stonnington.cynthia@mayo.edu.

INTRODUCTION

Even though simple blood tests [1–3] promise to become a useful and less expensive tool for assessing a person's diagnosis and prognosis in the early clinical and preclinical stages, structural magnetic resonance imaging (MRI) remains the most common biomarker assessment tool in current clinical practice. Typically, volumetric methods have utilized mass-univariate or region of interest methods to detect cortical thickness, grey matter volume, and surface areas. Accordingly, we recently reported findings from a prospective cohort of cognitively unimpaired individuals to estimate *a priori* MRI regions of interests (preferentially affected in the preclinical and clinical stages of Alzheimer's disease (AD)) for differences between those individuals who subsequently progressed to clinically significant memory decline in approximately 2 years and those who did not. Additionally, the same study also used Statistical Parametric Mapping (SPM) (http://www.fil.ion.ucl.ac.uk/spm/) to examine the ^{18}F-fluorodeoxyglucose (FDG) positron emission tomography (PET) measured cerebral metabolic rate for glucose (CMRgl) differences between progressors and nonprogressors. Based on receiver operating characteristic, binary logistic regression, and leave-one-out procedures, hippocampal volume best predicted an individual's imminent progression to the clinically significant memory decline, with 79% sensitivity/78% specificity among the *APOE*-matched cohort [4].

Multivariate methods appear to improve detection of subtle changes in MRI-based morphological features of structures relevant to preclinical detection of AD [5–7]. Machine learning methods promise to improve the accuracy of prediction for individual patients, particularly when applied to MRI based multiple features as with multivariate morphometry statistics (MMS), in the preclinical stages of AD [8–10]. In this study, we aimed to improve prediction from prior studies by employing the hippocampal surface MMS features, which have been shown to outperform the hippocampal volume measure [6, 11], and the patch based sparse coding algorithm to predict clinically significant memory impairment within two years, even in the absence of other amyloid, tau, PET, cerebrospinal fluid (CSF), or emerging blood-based biomarkers.

METHODS

Participants

Arizona cohort

These study participants were a sub-cohort of 280 drawn from our 23-year longitudinal Arizona *APOE* cohort study [12, 13]. As previously described [4], 18 "progressor" participants developed clinically significant memory impairment (16 diagnosed with amnestic MCI (aMCI), 1 with both amnestic and visuospatial MCI, and 1 with AD) and had both MRI and FDG PET data while still cognitively unimpaired at the epoch approximately 2 years prior to progression to aMCI/AD, and 20 "nonprogressor" participants who remained cognitively unimpaired at least 4 years after their last visits, all based on clinical, informant, neuropsychological data, and a Mini-Mental State Examination (MMSE) score > 26. The progressors and nonprogressors were matched for sex, age, education, and *APOE* allele dose. Participants with one abnormal score could be deemed clinically unimpaired if all other scores within the same cognitive domain were solidly normal and there were no functional impairments. The aMCI diagnosis was determined based on published criteria [14, 15] using clinical, functional, and neuropsychological data that included a wide battery of tests with > 1 test per domain. Though we subsequently introduced amyloid PET and tau PET to this overall study, we did not have these data for this specific sub-cohort available at baseline and therefore did not have amyloid or tau biomarkers to confirm AD pathology.

ADNI cohort

The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California – San Francisco. For up-to-date information, see http://www.adni-info.org. Study participants were drawn from ADNI data bases utilizing the same criteria to categorize and match progressors and nonprogressors described above for the Arizona cohort. From ADNI-1, ADNI-2, ADNI-Go, and ADNI-3 we found 60 participants who developed clinically significant memory impairment, i.e., aMCI, in approximately 2 years and 60 age, sex, education and *APOE*-matched nonprogressors who remained cognitively unimpaired for at least 4 years. "Baseline" scans were the MRI scans from progressors at 2 years prior to clinically significant decline and the corresponding matched nonprogressors' MRI scans.

The parent study for the Arizona cohort was approved by the Mayo Clinic and Banner Health (originally Banner Good Samaritan) Institutional Review Boards, and after complete description of the study to the subjects, written informed consent was obtained.

Hippocampus segmentation and surface reconstruction

All the T1-weighted MR images were automatically segmented by using FIRST [16], which is a model based subcortical structure registration and segmentation tool that we have used in our previous hippocampal morphometry research [6, 10, 11]. In comparison to FreeSurfer, FIRST is capable of generating topologically sound segmentation results with classification of relatively large, scaled databases. FIRST is one part of FSL library developed mainly by Analysis Group, FMRIB, Oxford, UK. With default parameters, we ran the *run_first_all* command and extracted the segmentation of left and right hippocampi. Then all the extracted images were binarized with a simple thresholding process. With the binary images, hippocampal surfaces were constructed with a topology-preserving level set method [18] and triangular surface meshes were further acquired based on marching cubes algorithm [19]. We then refined the meshes [6] to get the smooth surfaces which are suitable for generating conformal grids. Finally, all these smoothed meshes were aligned into the MNI standard space with a 9-DOF (degree of freedom) global affine transformation (Fig. 1).

Surface conformal representation

On each hippocampal surface, we generated a conformal grid as a canonical space for surface registration and multivariate statistical analysis [6]. Firstly, two cuts were introduced on the hippocampal surface (Fig. 2) and thus the surface could be converted into a tube-like genus zero surface. The two cuts locate at the front and back of the hippocampal surfaces, representing anterior junction with the amygdala, and its posterior limit as it turns into the white matter of the fornix. Thus, they are biologically valid and can be used as consistent landmarks across subjects. With the tube-like surface, the landmark curves can be automatically determined by locating the extreme points and searching along the first principle direction of geometric moments of the surface [7, 20, 21]. Finally, we calculated the holomorphic 1-form basis of each tube-like surface and conformally mapped the hippocampal surface to a planar surface.

Many of the geometric features of the surface could be contained in the conformal parameterization. In this paper, we calculated the local conformal factor and mean curvature, which represents the intrinsic and extrinsic features of the surface respectively. The conformal factor is the area ratio of the infinitesimal region around the same point on the original hippocampal surface and the conformal planar surface. Mean curvature is an extrinsic measure of curvature which comes from differential geometry and can represent the flatness of the surface around a vertex. Both the conformal factor and mean curvature are local features defined on each vertex. The conformal factor and mean curvature are called the *surface conformal representation* because they can encode both intrinsic structure and 3D embedding information.

Hippocampal surface registrations

All the hippocampal surfaces need to be registered to a common template surface for morphometric analysis. We used the aforementioned features, surface conformal factor and mean curvature, to enforce surface correspondence. So, with the conformal parameterization, we converted the 3D surface to a 2D image registration problem. We applied the well-studied image fluid registration algorithm [22, 23] to induce a deformation flow in the parameter domain. To simulate fluid flow on the surfaces, we introduced the Navier-Sokes equation into surface space using a manifold version of the Laplacian and divergence operators [24, 25]. With an inverse consistent framework, we could optimize the surface registration by minimizing the sum of squared surface feature intensity differences between the deforming image and the template. Since both the conformal mapping and the inverse consistent framework generate diffeomorphic mappings, the mapping between the surfaces is diffeomorphic.

Surface multivariate morphometry statistics

Surface MMS consists of two different features: multivariate tensor-based morphometry (mTBM) [26] and radial distance analysis [27, 28]. The mTBM can measure the deformation within the surface while the radial distance can measure hippocampal size according to the surface normal direction.

The mTBM statistics measure local surface deformation and have demonstrated improved signal

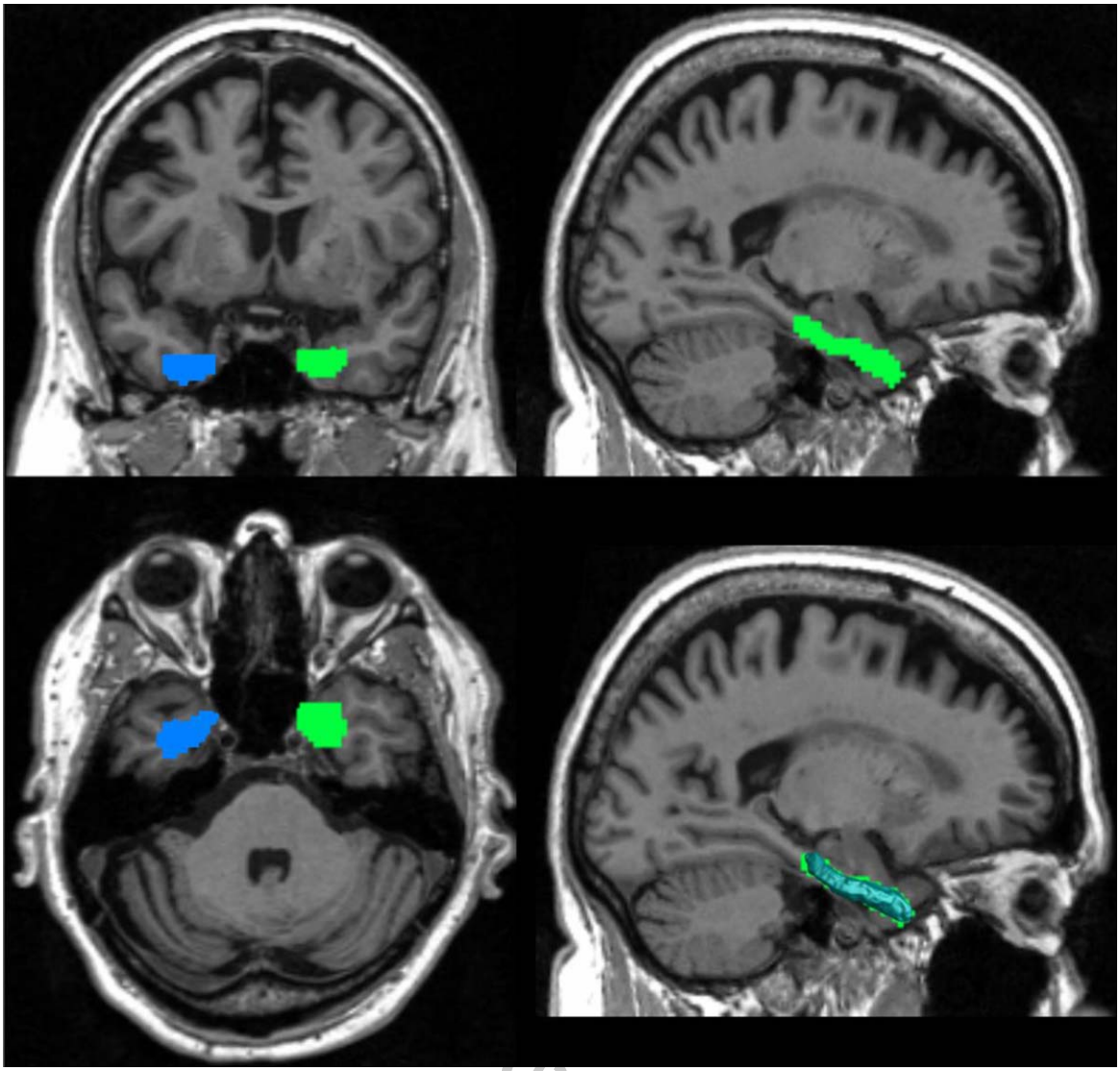

Fig. 1. Hippocampus segment The blue and green parts in these images represent the left and right hippocampus that FIRST segments of the image. The bottom right picture shows the shape of the hippocampal surface, which fits the segmented image well.

detection power relative to more standard tensor-based morphometry (TBM) measure computed as the determinant of Jacobian matrix [29].

Since the hippocampal surface is cut like a tube, the distance from each surface point to its medical core is affected by its atrophy and enlargement. We named the distance as the radial distance of a hippocampus surface, which represents the morphometric changes along the surface normal direction. Thus, radial distance and mTBM are complementary to each other; finally, we formed the new multivariate surface morphometry statistic as a 4×1 vector, of which the mTMB was computed as a 3×1 vector consisting of the "Log-Euclidean metric" [30] and the radial distance is just a scalar.

Patch analysis-based surface correntropy-induced- sparse-coding (PASCS)

Recently, sparse representation and sparse coding methodology developed in the machine learning field has been shown to be efficient in learning diverse and discriminative features for optimal representations [31, 32]. Our prior work adopting sparse coding for MRI data analysis in AD showed promising performance [33–36]. The basic idea of sparse coding is

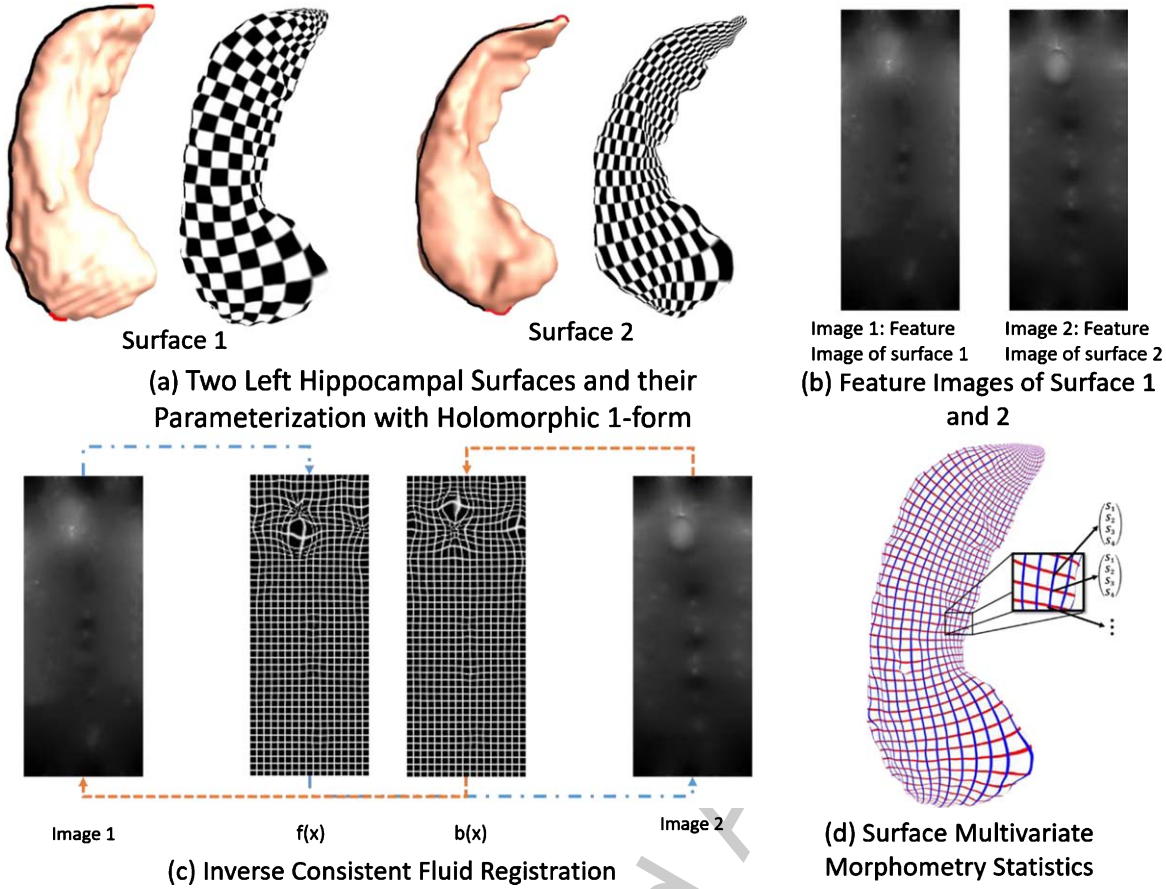

Fig. 2. Hippocampal surface morphometry pipeline (a) The hippocampus is segmented from T1-weighted images and a conformal grid is built on the surface. Here examples are shown for 2 different subjects. (b) Examples of features selected in the image for the two subjects in (a). (c) From left to right: Intensity map on the surface 1 in (a). Forward map f(x) for the conformal grid fluid registration to the image 2. Backward map b(x) from the image 2 to image 1. Intensity map on the surface 2. (d). Surface multivariate morphometry statistics is applied to analyze morphometric changes.

to generate an over-complete dictionary that allows us to represent the original high-dimensional features with a sparse coefficient matrix (sparse codes) for learning the optimal representation. The advantage of sparse coding is that it can use a small number of basis vectors to represent local features effectively and concisely and help extract the most discriminative features for image content analysis. Sparse coding has shown to be efficient for many medical image tasks, including image classification [37], image denoising [38], image segmentation [39], and functional connectivity [40]. In our research, we use the combination of surface patch features as input and construct both dictionary and their sparse codes to reconstruct the input features. Usually, the objective function aims to optimize two terms: the first term measures how well it represents the surface patches, and the second term ensures the sparsity of the representation, with an l_1-regularized correntropy loss function. In this work, stochastic coordinate coding [42] is adopted due to its ability to dramatically reduce the computational cost while keeping comparable performance. We further use the learned sparse representation as surface features.

Patch selection with sparse coding

After registering each hippocampal surface to a uniformed grid, each surface contains 150 * 100 vertices and the feature dimension of each hippocampal surface is 60,000, where each vertice has 1 * 4 dimensional MMS features. We then randomly generated 10×10 square windows on each hippocampal surface and collected 504 surface patches with different amounts of overlapping on each side of the hippocampus. We randomly selected 1008 patches on each

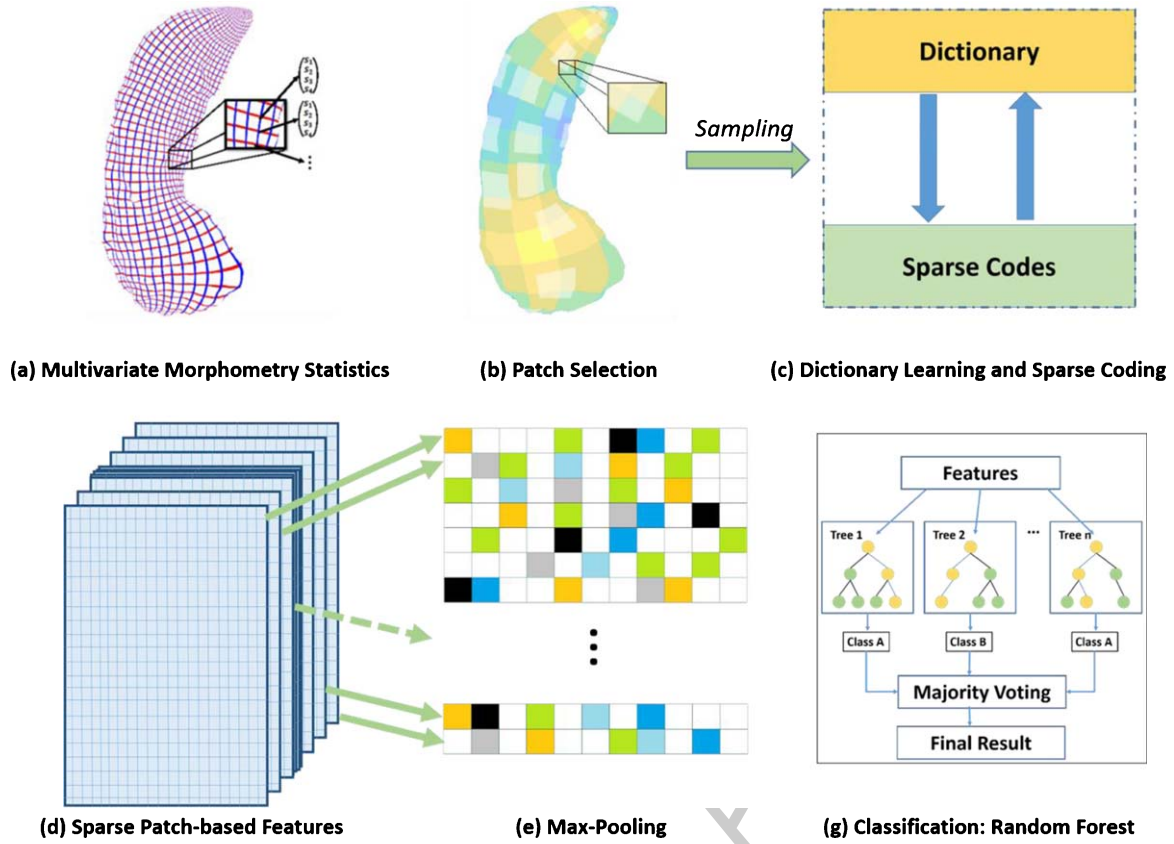

Fig. 3. Patch-based sparse coding system (a) Surface multivariate morphometry statistics. (b) Generate patches and randomly select patches on the surface. (c) Dictionary learning and sparse coding. (d) Sparse patch-based features got from (c). (e) Max-pooling to resize the features in (d). (g) Classification by using random forest classifier with the features after Max-pooling.

subject's hippocampal surfaces (1008 for both left and right). For different subjects, we used the same random seed to choose the patches. In other words, the distribution for the random-selected patches is the same on the hippocampal surfaces for all the subjects. Then we reformed these patches of features to a vector, of which the dimension is 400×1008. The dictionary was initialized by randomly selecting patches [43], which has proved to be an efficient method in practice, and then we started learning the dictionary and sparse codes by stochastic coordinate coding [42]. The size of the batch is one and the model is trained for ten epochs. After sparse coding, we acquired 1008 samples, each of which has 1800 features on each subject. Finally, with max-pooling, we chose the maximum values for each feature over 1008 patches and obtained 1800-dimensional features for each subject.

In this study, we chose random forest algorithm [44]. Random forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest (Fig. 3). This algorithm adapts a learning process called "feature bagging." In this process, we selected a random subset of the features for several times and then trained a decision tree for each subset. If some features are strong predictors for the response, they will be selected in many decision trees and thus make them correlated. In comparison with decision trees, random forests have the same bias but lower variance, which means it can overcome the drawback of overfitting caused by the small data set. For our sparse surface features, when the training number becomes smaller, diversification becomes more subtle, and the method can better detect these subtle differences. Finally, we employed cross-validation to evaluate the performance of the classification. For the k-fold cross-validation, we randomly shuffled the dataset and split it to k groups. For each group, we take it as the test data set and use the remaining groups to train a model. Then, the model

Table 1
Characteristics progressors and nonprogressors at the time of baseline scan

	Progressors	Nonprogressors	p
Sex (M/F)	7/11 (AZ)	7/13 (AZ)	0.80
	25/35 (ADNI)	25/35 (ADNI)	1.00
ε4 Genotype (N)	13:03:02 (AZ)	13:04:03 (AZ)	0.89
	2:17:41 (ADNI)	1:17:42 (ADNI)	0.99
% (HM:HT: NC)	(72:17:11) (AZ)	(65:20:15) (AZ)	
	(0:39:61) (ADNI)	(0:21:79) (ADNI)	
Age	68.75 ± 4.65 (AZ)	66.76 ± 3.29 (AZ)	0.13
	76.97 ± 6.89 (ADNI)	75.19 ± 5.62 (ADNI)	0.12
Education	16.44 ± 1.69 (AZ)	15.50 ± 3.33 (AZ)	0.29
	15.95 ± 2.87 (ADNI)	16.12 ± 2.74 (ADNI)	0.70

Sex and genotype p-values were calculated by chi-squared tests, Age and education p-values were calculated by t-tests. HM, ε4 homozygote; HT, ε4 heterozygote; NC, ε4 non-carrier; AZ, Arizona cohort; ADNI, Alzheimer's Disease Neuroimaging Initiative cohort.

is evaluated by the test group. In this way, we can get a predicted class label for all the samples. To indicate the number of correct class labels, we built a contingency table, of which the rows are the true classes and the columns represent assigned classes. And then, we could represent the combination of ground truth and predicted result as $\begin{Bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{Bmatrix}$ and compute the following performance measures, Sensitivity $= C_{11}/(C_{11} + C_{12})$, Specificity $= C_{22}/(C_{21} + C_{22})$ and Accuracy $= (C_{11} + C_{22})/(C_{11} + C_{12} + C_{21} + C_{22})$.

Data availability

Any data not published within the article is available, and anonymized data will be shared by request from any qualified investigator.

RESULTS

Characteristics of the progressors and nonprogressors are shown in Table 1. Overall, the ADNI participants were older and had a greater percentage of males and *APOE*4 non-carriers than the Arizona participants.

Prediction results are shown in Tables 2 and 3. In the Arizona cohort, the prediction result of progression to clinically significant decline using hippocampal surface MMS features was achieved with 92% accuracy and 89% sensitivity and 95% specificity. The same method with the ADNI cohort achieved 92% accuracy, 88% sensitivity, and 97% specificity. Combining the Arizona and ADNI cohorts (78 progressors and 80 nonprogressors) achieved 90% prediction accuracy, 90% sensitivity, and 90% specificity.

A *post-hoc* analysis using random forest classification and leave-one-out cross-validation showed that AVLT long-term memory (LTM) scores predicted progression with only 74% prediction accuracy (65% sensitivity and 83% specificity) in the Arizona cohort, and 62% prediction accuracy (59% sensitivity and 65% specificity) in the ADNI cohort despite the potential bias of using that same measure (along with other criteria) when making the diagnosis of aMCI. Furthermore, as a comparison to our new methods that utilize surface multivariate morphometry, prediction of aMCI using baseline hippocampal volume, random forest classification with leave-one-out cross-validation in the same data sets yielded only 79% prediction accuracy in the Arizona cohort and 72% in the ADNI cohort (Table 4).

Table 2
Experimental results: Arizona and ADNI cohorts (leave-one-out cross-validation)

Hippocampal surface	MMS	MMS left	MMS right
Accuracy	**0.92 (AZ)**	0.74 (AZ)	0.66 (AZ)
	0.92 (ADNI)	0.85 (ADNI)	0.84 (ADNI)
Sensitivity	**0.89 (AZ)**	0.72 (AZ)	0.61 (AZ)
	0.88 (ADNI)	0.84 (ADNI)	0.83 (ADNI)
Specificity	**0.95 (AZ)**	0.75 (AZ)	0.70 (AZ)
	0.97 (ADNI)	0.87 (ADNI)	0.85 (ADNI)

Multivariate morphometry statistics (MMS) column indicates the classification results with MMS from both left and right hippocampal surfaces while MMS left and MMS right columns are the classification results with MMS from left and right hippocampal surfaces, respectively. AZ, Arizona Cohort; ADNI, Alzheimer's Disease Neuroimaging Initiative Cohort.

Table 3
Experimental results: combined cohorts (5-fold cross-validation)

Hippocampal surface	MMS	MMS left	MMS right
Accuracy	**0.90**	0.84	0.80
Sensitivity	**0.90**	0.82	0.79
Specificity	**0.90**	0.85	0.81

Multivariate morphometry statistics (MMS) column indicates the classification results with MMS from both left and right hippocampal surfaces while MMS left and MMS right columns are the classification results with MMS from left and right hippocampal surfaces, respectively.

DISCUSSION

This study extended previous work by showing that combining hippocampal surface MMS and

Table 4
Hippocampal volume prediction using random forest classifier and leave-one-out cross-validation

Hippocampal Volume*	Accuracy	Sensitivity	Specificity
Left hippocampus	0.74 (AZ)	0.65 (AZ)	0.83 (AZ)
	0.68 (ADNI)	0.68 (ADNI)	0.68 (ADNI)
Right hippocampus	0.74 (AZ)	0.80 (AZ)	0.67 (AZ)
	0.65 (ADNI)	0.60 (ADNI)	0.70 (ADNI)
Left+Right hippocampus	**0.79 (AZ)**	0.85 (AZ)	0.72 (AZ)
	0.72 (ADNI)	0.69 (ADNI)	0.73 (ADNI)

*The automated brain mapping algorithmic program FreeSurfer (http://surfer.nmr.mgh.harvard.edu/) was used to pre-process MRI volumetric data.

machine learning methods affords improved prediction of imminent clinically significant cognitive decline compared to typical automated volumetric MRI methods and standard statistics. Our methods achieved high prediction accuracy in two separate and independent data sets, each balanced for age, sex, and presence or absence of the *APOE4* allele, even in the absence of other brain imaging or fluid biomarkers of AD. Furthermore, we retained accuracy when the two data sets, which differed from each other in age, sex, and percentage *APOE4* carriers, were combined. The predictions using hippocampal surface MMS and machine learning methods were also much better than predictions using either hippocampal volume or baseline cognitive scores, even though the latter are biased due to the circularity of using the same measure when making the diagnosis of aMCI. Although lacking amyloid and tau biomarker confirmation of AD pathology in this study, the data set from Arizona is well-defined, with high confidence regarding the likelihood of AD in those who subsequently developed aMCI. Thus far, the majority of those in the Arizona progressor group later developed definite or probable AD, with the exception of one person who developed dementia with Lewy bodies and one aMCI individual who 2 years later had slight improvement in cognition. Although we do not know how many of those in the nonprogressor group will ultimately develop AD, we have high confidence that none developed MCI for 4 years following the scan. To date, 3 in the Arizona nonprogressor group subsequently developed MCI and none have progressed to dementia. Thus, the training and testing groups were well defined and mostly accurate, which mirrors the accuracy of our novel feature-based sparse coding methods. Instead of using the automated brain mapping algorithmic programs FreeSurfer (http://surfer.nmr.mgh.harvard.edu/) and Statistical Parametric Mapping (http://www.fil.ion.ucl.ac.uk/spm/) as we did in our prior study with the same data set [4], in this study we utilized arguably more sensitive methods involving MMS to discover subregional hippocampal surface differences, patched-based sparse coding for feature selection, and the random forest machine learning classifier. We were able to replicate our improved results in a completely independent data set from ADNI that differed from the Arizona data set in age, sex, and percentage *APOE*4 carriers.

Because the Arizona data set was identical to our prior study [4], the improved accuracy in this study can be explained by our use of hippocampal surface MMS combined with patch-based sparse coding algorithms. Similar to the methods from our most recent work [36], in this paper we propose a novel Patch Analysis-based Surface Correntropy-induced Sparse coding, PASCS, to help predict future cognitive decline. We demonstrate that PASCS is surprisingly useful for surface features classification and surface multivariate morphometry statistics features consisting of surface multivariate tensor-based morphometry and radial distance (the distance from the medial core to each surface point), and we also move a step forward from group difference to that of individual subject classification. Unlike other sparse coding work [37–40], PASCS takes advantage of surface morphometry features that practically encode neighboring intrinsic 3D geometry information. Meanwhile, MMS features also benefit from the succinct representation and strong discrimination power that sparse coding provides for effective AD classifications, i.e., capturing more important information so that MMS features not only have the significant group difference but also have an effective classification power.

In this work, we adopted FIRST for hippocampus segmentation, which, having previously explored different segmented hippocampal data as input, appears to most reliably generate topologically sound segmentation results. For example, our earlier work used manually segmented hippocampi to build surface meshes [26, 45]. Later, we adopted FIRST for automatic hippocampus segmentation [6] and used it in almost all our hippocampal morphometry research. Meanwhile, we also used FreeSurfer segmented hippocampi to build hippocampal surface meshes [17]. All achieved reasonable results in group difference studies, thus demonstrating that our pipeline is robust to segmentation methods. However, FIRST can always generate topologically sound segmentation results, whereas FreeSurfer does not guarantee

Table 5
Classifications for 10 different patch selections

Cohorts	AZ	ADNI	Combined
Accuracy ± SD	0.89 ± 0.02	0.90 ± 0.01	0.90 ± 0.02
Sensitivity ± SD	0.85 ± 0.04	0.89 ± 0.02	0.90 ± 0.04
Specificity ± SD	0.93 ± 0.06	0.91 ± 0.02	0.90 ± 0.04

We repeated the experiments ten times with ten different patch selections on both sides of the hippocampal surface. The left and middle columns of the table indicate the classification results for the AZ cohort and ADNI cohort respectively with leave-one-out cross-validation. The last column shows the results for the combined cohorts with 5-fold cross-validation. SD, standard deviation.

topologically correct results. Therefore, manual quality control is necessary to incorporate FreeSurfer in our pipeline. Thus far, our related prediction/classification work adopted FIRST segmented hippocampal surfaces in order to more efficiently work with relatively large-scaled datasets [10, 11]. Since the input of our MMLC is the surface features rather than the output from segmentation tools, it is reasonable for us to expect that our method is not sensitive to the hippocampus segmentation tools used.

To evaluate the influence of these random-select patches on the classification accuracy and the stability of our framework, we repeated the experiments ten times with ten different patch selection on both sides of the hippocampal surface. The mean and standard deviation of the results are shown in Table 5. The results show our method is relatively stable with different patch selection and comparable to the best accuracy results reported in Tables 2 and 3. It is worth noting that the variance results were not purely caused by the patch selection, since other components in the pipeline, such as random forest and cross-validation parts, may also perturb the final results. For example, during the training of random forest, the classifier will randomly select a subset of features to build a decision tree. Similarly, in the 5-fold cross-validation, the training data may be different. Considering the small dataset size in the current experiments, the minor variance in our results demonstrates that the influence of random patch selection is in a reasonable range and does not appreciably affect the stability of the results. In future, we will further explore the random patch selection issue with larger imaging cohorts.

Future directions will include integrating convolutional neural network (CNN) with our proposed approach. CNN is considered one of the most successful deep models for identifying, classifying, and quantifying patterns in medical images [46, 47]. Based on promising results from our most recent work [48, 49], integrating CNN with the proposed approach could further improve the PASS results. Specifically, we applied CNN and an unsupervised learning method (multi-task stochastic coordinate coding) algorithm to the ADNI dataset to predict future cognitive clinical measures with baseline hippocampal/ventricle mTBM features and cortical thickness, achieving accurate predictions of MMSE/ADAS-Cog scales [48, 49]. However, there is a trade-off between computation efficiency and prediction performance. Training a CNN model usually requires substantial computational resources (multiple GPUs). Our PASS-MP is a generative toolbox for brain image analysis with fast running time and does not require GPU for training. It can apply to different subcortical of brain images with relatively high performance. We therefore will continue to explore the efficient CNN based sparse coding method that could both improve the prediction power and maintain a low-cost of computational resources and fast running time as PASS-MP for better help with clinical diagnosis and prognosis.

Although there are many other sensitive biomarkers to detect the pathology associated with AD, this method capitalizes on MRI scans, which is a clinical diagnostic capability that virtually all clinicians have access to. Further testing is needed to verify the results in larger data sets, but our method appears to accurately predict whether an individual will progress to the clinical stages of AD within the next 2 years. Thus, this method has the potential to be developed into a clinically useful tool. We currently have no proven medication treatments for AD; however, well-tested behavioral programs that provide lifestyle and behavioral training to adapt to memory loss associated with MCI are available [50]. These behavioral programs appear to be most effective when done prior to significant memory decline [51]. If we had an accurate and inexpensive tool to predict likelihood of clinically significant decline, we could target those individuals who would most benefit from a similar intervention that is delivered preclinically.

Limitations of this study include relatively small numbers of progressors and nonprogressors in both cohorts and thus this method will need to be replicated in other, larger data sets. Importantly, we also did not include other biomarkers such as amyloid or tau to verify that progressors had MCI due to AD or include other imaging, CSF, and emerging, less expensive and more scalable blood-based biomarkers of amyloid, tau, and neurodegeneration [1–3], and future applications of this technique should do

so to ensure accurate training sets and generalizable results. However, we were interested in seeing the "added value" of this MRI based image analysis technique as a complement to those emerging methods. We did not test brain regions other than hippocampus, but the purpose of this study was to evaluate the utility of prediction using hippocampal surface multivariate morphometry statistics combined with patch-based sparse coding algorithms. It was therefore convenient to compare these methods using the same data set that we had previously evaluated with standard automated brain mapping algorithmic programs, binary logistic regression, and leave-one-out procedures. Also, our previous study did explore other imaging-based biomarkers and found that the hippocampus was the best predictor 2 years prior to clinically significant decline (including both FDG-PET and MRI biomarkers). Finally, because of the overlapping patch selection and max-pooling scheme, we generally cannot visualize the selected features, which may decrease the interpretability of biomarkers and, in turn, translation to clinical applications. However, we can always visualize statistically significant regions using group differences [6]. In addition, our recent work [52] better addresses this problem with the adoption of group lasso screening [53] to select the most significant features. It was not adopted in our current study because of its relatively small sample size. In the future, we will incorporate this approach into our current framework to improve its interpretability.

ACKNOWLEDGMENTS

The authors thank Bruce Henslin and Kathryn DeMarco for their help with data management. This work was supported by NIH APOE4 grant, "Brain Imaging, APOE and the Preclinical Course of Alzheimer's Disease" (R01AG0311581); ADCC grant "Alzheimer's Disease Core Center" (P30AG 019610); and the State of Arizona. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

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-Myers Squibb 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 (http://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.

Authors' disclosures available online (https://www.j-alz.com/manuscript-disclosures/20-0821r2).

REFERENCES

[1] Bateman RJ, Blennow K, Doody R, Hendrix S, Lovestone S, Salloway S, Schindler R, Weiner M, Zetterberg H, Aisen P, Vellas B (2019) Plasma piomarkers of AD emerging as essential tools for drug development: An EU/US CTAD Task Force Report. *J Prev Alzheimers Dis* **6**, 169-173.

[2] Janelidze S, Mattsson N, Palmqvist S, Smith R, Beach TG, Serrano GE, Chai X, Proctor NK, Eichenlaub U, Zetterberg H, Blennow K, Reiman EM, Stomrud E, Dage JL, Hansson O (2020) Plasma P-tau181 in Alzheimer's disease: Relationship to other biomarkers, differential diagnosis, neuropathology and longitudinal progression to Alzheimer's dementia. *Nat Med* **26**, 379-386.

[3] Palmqvist S, Insel PS, Stomrud E, Janelidze S, Zetterberg H, Brix B, Eichenlaub U, Dage JL, Chai X, Blennow K, Mattsson N, Hansson O (2019) Cerebrospinal fluid and plasma biomarker trajectories with increasing amyloid deposition in Alzheimer's disease. *EMBO Mol Med* **11**, e11170.

[4] Stonnington CM, Chen Y, Savage CR, Lee W, Bauer Iii RJ, Sharieff S, Thiyyagura P, Alexander GE, Caselli RJ, Locke DEC, Reiman EM, Chen K (2018) Predicting imminent progression to clinically significant memory decline using volumetric MRI and FDG PET. *J Alzheimers Dis* **63**, 603-615.

[5] Li S, Yuan X, Pu F, Li D, Fan Y, Wu L, Chao W, Chen N, He Y, Han Y (2014) Abnormal changes of multidimensional surface features using multivariate pattern classification in amnestic mild cognitive impairment patients. *J Neurosci* **34**, 10541-10553.

[6] Shi J, Thompson PM, Gutman B, Wang Y, Alzheimer's Disease Neuroimaging Initiative (2013) Surface fluid registration of conformal representation: Application to detect disease burden and genetic influence on hippocampus. *Neuroimage* **78**, 111-134.

[7] Wang Y, Song Y, Rajagopalan P, An T, Liu K, Chou YY, Gutman B, Toga AW, Thompson PM (2011) Surface-based TBM boosts power to detect disease effects on the brain: An *N*=804 ADNI study. *Neuroimage* **56**, 1993-2010.

[8] Rathore S, Habes M, Iftikhar MA, Shacklett A, Davatzikos C (2017) A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages. *Neuroimage* **155**, 530-548.

[9] Zhang J, Li Q, Caselli RJ, Thompson PM, Ye J, Wang Y (2017) Multi-source multi-target dictionary learning for prediction of cognitive decline. *Inf Process Med Imaging* **10265**, 184-197.

[10] Fu Y, Zhang J, Li Y, Shi J, Zou Y, Guo H, Li Y, Yao Z, Wang Y, Hu B (2020) A novel pipeline leveraging surface-based features of small subcortical structures to classify individuals with autism spectrum disorder. *Prog Neuropsychopharmacol Biol Psychiatry* **104**, 109989.

[11] Dong Q, Zhang W, Wu J, Li B, Schron EH, McMahon T, Shi J, Gutman BA, Chen K, Baxter LC, Thompson PM, Reiman EM, Caselli RJ, Wang Y (2019) Applying surface-based hippocampal morphometry to study APOE-E4 allele dose effects in cognitively unimpaired subjects. *Neuroimage Clin* **22**, 101744.

[12] Caselli RJ, Dueck AC, Osborne D, Sabbagh MN, Connor DJ, Ahern GL, Baxter LC, Rapcsak SZ, Shi J, Woodruff BK, Locke DE, Snyder CH, Alexander GE, Rademakers R, Reiman EM (2009) Longitudinal modeling of age-related memory decline and the APOE epsilon4 effect. *N Engl J Med* **361**, 255-263.

[13] Caselli RJ, Reiman EM, Osborne D, Hentz JG, Baxter LC, Hernandez JL, Alexander GG (2004) Longitudinal changes in cognition and behavior in asymptomatic carriers of the APOE e4 allele. *Neurology* **62**, 1990-1995.

[14] Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, Gamst A, Holtzman DM, Jagust WJ, Petersen RC, Snyder PJ, Carrillo MC, Thies B, Phelps CH (2011) The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* **7**, 270-279.

[15] McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Jr., Kawas CH, Klunk WE, Koroshetz WJ, Manly JJ, Mayeux R, Mohs RC, Morris JC, Rossor MN, Scheltens P, Carrillo MC, Thies B, Weintraub S, Phelps CH (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. *Alzheimers Dement* **7**, 263-269.

[16] Patenaude B, Smith SM, Kennedy DN, Jenkinson M (2011) A Bayesian model of shape and appearance for subcortical brain segmentation. *Neuroimage* **56**, 907-922.

[17] Joshi SH, Espinoza RT, Pirnia T, Shi J, Wang Y, Ayers B, Leaver A, Woods RP, Narr KL (2016) Structural plasticity of the hippocampus and amygdala induced by electroconvulsive therapy in major depression. *Biol Psychiatry* **79**, 282-292.

[18] Han X, Xu CY, Prince JL (2003) A topology preserving level set method for geometric deformable models. *IEEE Trans Pattern Anal Mach Intell* **25**, 755-768.

[19] Lorensen WE, Cline HE (1987) Marching cubes: A high resolution 3D surface construction algorithm. *SIGGRAPH Comput Graph* **21**, 163-169.

[20] Elad M, Milanfar P, Golub GH (2004) Shape from moments - An estimation theory perspective. *IEEE Trans Signal Process* **52**, 1814-1829.

[21] Zhang DS, Lu GJ (2004) Review of shape representation and description techniques. *Pattern Recognit* **37**, 1-19.

[22] Bro-Nielsen M, Gramkow C (1996) *Visualization in Biomedical Computing (VBC'96)* Springer.

[23] D'Agostino E, Maes F, Vandermeulen D, Suetens P (2003) A viscous fluid model for multimodal non-rigid image registration using mutual information. *Med Image Anal* **7**, 565-575.

[24] Aris R (1989) *Vectors, tensors, and the basic equations of fluid mechanics*, Dover Publications, New York.

[25] Stam J (2003) Flows on surfaces of arbitrary topology. *Proceedings of SIGGRAPH*, pp. 723-731.

[26] Wang Y, Chan TF, Toga AW, Thompson PM (2009) Multivariate tensor-based brain anatomical surface morphometry via holomorphic one-forms. *Med Image Comput Comput Assist Interv* **12**, 337-344.

[27] Pizer SM, Fritsch DS, Yushkevich PA, Johnson VE, Chaney EL (1999) Segmentation, registration, and measurement of shape variation via image object shape. *IEEE Trans Med Imaging* **18**, 851-865.

[28] Thompson PM, Hayashi KM, De Zubicaray GI, Janke AL, Rose SE, Semple J, Hong MS, Herman DH, Gravano D, Doddrell DM, Toga AW (2004) Mapping hippocampal and ventricular change in Alzheimer disease. *Neuroimage* **22**, 1754-1766.

[29] Chung MK, Dalton KM, Davidson RL (2008) Tensor-based cortical surface morphometry via weighted spherical harmonic representation. *IEEE Trans Med Imaging* **27**, 1143-1151.

[30] Arsigny V, Fillard P, Pennec X, Ayache N (2006) Log-Euclidean metrics for fast and simple calculus on diffusion tensors. *Magn Reson Med* **56**, 411-421.

[31] Mairal J, Bach F, Ponce J, Sapiro G (2010) Online learning for matrix factorization and sparse coding. *J Mach Learn Res* **11**, 19-60.

[32] Vu TH, Monga V (2017) Fast low-rank shared dictionary learning for image classification. *IEEE Trans Image Process* **26**, 5160-5175.

[33] Zhang J, Fan Y, Li Q, Thompson PM, Ye J, Wang Y (2017) Empowering cortical thickness measures in clinical diagnosis of Alzheimer's disease with spherical sparse coding. *Proc IEEE Int Symp Biomed Imaging* **2017**, 446-450.

[34] Zhang J, Shi J, Stonnington C, Li Q, Gutman BA, Chen K, Reiman EM, Caselli RJ, Thompson PM, Ye J, Wang Y (2016) Hyperbolic space sparse coding with its application on prediction of Alzheimer's disease in mild cognitive impairment. *Med Image Comput Comput Assist Interv* **9900**, 326-334.

[35] Zhang J, Stonnington C, Li Q, Shi J, Bauer RJ, 3rd, Gutman BA, Chen K, Reiman EM, Thompson PM, Ye J, Wang Y (2016) Applying sparse coding to surface multivariate

tensor-based morphometry to predict future cognitive decline. *Proc IEEE Int Symp Biomed Imaging* **2016**, 646-650.

[36] Wu J, Dong Q, Gui J, Zhang J, Su Y, Chen K, Thompson PM, Caselli RJ, Reiman EM, Ye J, Wang Y (2020) Predicting brain amyloid using multivariate morphometry statistics, sparse coding, and correntropy: Validation in 1,125 individuals from the ADNI and OASIS databases. *bioRxiv*, 2020.2010.2016.343137.

[37] Shi J, Li Y, Zhu J, Sun H, Cai Y (2015) Joint sparse coding based spatial pyramid matching for classification of color medical image. *Comput Med Imaging Graph* **41**, 61-66.

[38] Staglianò A, Chiusano G, Basso C, Santoro M (2010) Learning adaptive and sparse representations of medical images In *International MICCAI Workshop on Medical Computer Vision* Springer, Berlin, Heidelberg, pp. 130-140.

[39] Zhang ST, Zhan YQ, Metaxas DN (2012) Deformable segmentation via sparse representation and dictionary learning. *Med Image Anal* **16**, 1385-1396.

[40] Lv JL, Ling BB, Li QY, Zhang W, Zhao Y, Jiang X, Guo L, Han JW, Hu XT, Guo C, Ye JP, Liu TM (2017) Task fMRI data analysis based on supervised stochastic coordinate coding. *Med Image Anal* **38**, 1-16.

[41] Fu WJ (1998) Penalized regressions: The bridge versus the lasso. *J Comput Graph Stat* **7**, 397-416.

[42] Lin B, Li Q, Sun Q, Lai MJ, Davidson I, Fan W, Ye J (2014) Stochastic coordinate coding and its application for Drosophila gene expression pattern annotation. *arXiv:1407.8147v2*

[43] Coates A, Ng AY (2011) The importance of encoding versus training with sparse coding and vector quantization. *Proceedings of the 28th International Conference on Machine Learning* **11**, 921-928.

[44] Breiman L (2001) Random forests. *Mach Learn* **45**, 5-32.

[45] Luders E, Thompson PM, Kurth F, Hong JY, Phillips OR, Wang Y, Gutman BA, Chou YY, Narr KL, Toga AW (2013) Global and regional alterations of hippocampal anatomy in long-term meditation practitioners. *Hum Brain Mapp* **34**, 3369-3375.

[46] Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak JAWM, van Ginneken B, Sanchez CI (2017) A survey on deep learning in medical image analysis. *Med Image Anal* **42**, 60-88.

[47] Shen DG, Wu GR, Suk HI (2017) Deep learning in medical image analysis. *Annu Rev Biomed Eng* **19**, 221-248.

[48] Dong Q, Zhang J, Li Q, Thompson PM, Caselli RJ, Ye J, Wang Y, Initiative AsDN (2019) Multi-task dictionary learning based on convolutional neural networks for longitudinal clinical score predictions in Alzheimer's disease. In *International Workshop on Human Brain and Artificial Intelligence*, Zeng A, Pan D, Hao T, Zhang D, Shi Y, Song X, eds. Springer Singapore, pp. 21-35.

[49] Dong Q, Zhang J, Li Q, Wang J, Lepore N, Thompson PM, Caselli RJ, Ye J, Wang Y (2020) Integrating convolutional neural networks and multi-task dictionary learning for cognitive decline prediction with longitudinal images. *J Alzheimers Dis* **75**, 971-992.

[50] Chandler MJ, Locke DE, Crook JE, Fields JA, Ball CT, Phatak VS, Dean PM, Morris M, Smith GE (2019) Comparative effectiveness of behavioral interventions on quality of life for older adults with mild cognitive impairment: A randomized clinical trial. *JAMA Netw Open* **2**, e193016.

[51] De Wit L, Chandler M, Amofa P, DeFeis B, Mejia A, O'Shea D, Locke DEC, Fields JA, Smith GE (2021) Memory Support System training in mild cognitive impairment: Predictors of learning and adherence. *Neuropsychol Rehabil* **31**, 92-104.

[52] Zhang J, Tu Y, Li Q, Caselli RJ, Thompson PM, Ye J, Wang Y (2018) Multi-task sparse screening for predicting future clinical scores using longitudinal cortical thickness measures. *Proc IEEE Int Symp Biomed Imaging* **2018**, 1406-1410.

[53] Wang J, Wonka P, Ye JP (2015) Lasso screening rules via dual polytope projection. *J Mach Learn Res* **16**, 1063-1101.