Diagnosis and Prognosis Using Machine Learning Trained on Brain Morphometry and White Matter Connectomes

Yun Wang[1], Chenxiao Xu[2], Ji-Hwan Park[2], Seonjoo Lee[3],
Yaakov Stern[4], Shinjae Yoo[5], Jong Hun Kim[6], Hyoung Seop Kim[7], Jiook Cha[1]

1. Department of Psychiatry, Columbia University Medical Center, New York, NY, USA;
2. Department of Applied Mathematics, Stony Brook University, Stony Brook, NY, USA;
3. Department of Biostatistics, School of Public Health, Columbia University Medical Center, New York, NY, USA;
4. Department of Neurology, Columbia University Medical Center, New York, NY, USA;
5. Computational Science Initiative, Brookhaven National Laboratory, Upton, NY, USA;
6. Department of Neurology, National Health Insurance Service Ilsan Hospital, Goyang, Republic of Korea
7. Department of Physical Medicine and Rehabilitation, National Health Insurance Service Ilsan Hospital, Goyang, Republic of Korea

Corresponding to:

Jiook Cha, PhD
1051 Riverside Dr. New York, NY, 10032
Jiook.cha@nyspi.columbia.edu

Hyoung Seop Kim, MD
100 Ilsan-ro, Ilsandong-gu, Goyang-si, Gyeonggi-do,10444 South Korea
rehappydoc@gmail.com

ABSTRACT

Accurate, reliable prediction of risk for Alzheimer's disease (AD) is essential for early, disease-modifying therapeutics. Multimodal MRI, such as structural and diffusion MRI, is likely to contain complementary information of neurodegenerative processes in AD. Here we tested the utility of commonly available multimodal MRI (T1-weighted structure and diffusion MRI), combined with high-throughput brain phenotyping—morphometry and connectomics—and machine learning, as a diagnostic tool for AD. We used, firstly, a clinical cohort at a dementia clinic (study 1: Ilsan Dementia Cohort; N=211; 110 AD, 64 mild cognitive impairment [MCI], and 37 subjective memory complaints [SMC]) to test and validate the diagnostic models; and, secondly, Alzheimer's Disease Neuroimaging Initiative (ADNI)-2 (study 2) to test the generalizability of the approach and the prognostic models with longitudinal follow up data. Our machine learning models trained on the morphometric and connectome estimates (number of features=34,646) showed optimal classification accuracy (AD/SMC: 97% accuracy, MCI/SMC: 83% accuracy; AD/MCI: 97% accuracy) with iterative nested cross-validation in a single-site study, outperforming the benchmark model (FLAIR-based white matter hyperintensity volumes). In a generalizability study using ADNI-2, the combined connectome and morphometry model showed similar or superior accuracies (AD/HC: 96%; MCI/HC: 70%; AD/MCI: 75% accuracy) as CSF biomarker model (t-tau, p-tau, and Amyloid β, and ratios). We also predicted MCI to AD progression with 69% accuracy, compared with the 70% accuracy using CSF biomarker model. The optimal classification accuracy in a single-site dataset and the reproduced results in multi-site dataset show the feasibility of the high-throughput imaging analysis of multimodal MRI and data-driven machine learning for predictive modeling in AD.

Keywords: Alzheimer's disease; Multimodal MRI; White Matter Structural Connectomes; Probabilistic Tractography; Machine Learning

INTRODUCTION

There is an urgent, unmet need for clinically useful biomarkers of risk for Alzheimer's disease (AD) based on non-invasive and affordable measures suited for routine examination of individuals with subthreshold symptoms. Studies have focused on brain MRI-derived markers. Cortical thinning and reduced hippocampal volumes based on structural MRI are known for markers for AD, but these structural estimates alone are insufficient for implementation at clinical settings because of insufficient accuracy and generalizability (Teipel et al., 2015).

It is conceptualized that biomarkers of Aβ deposition become abnormal early, and then markers of neuronal neurodegeneration or dysfunction show abnormality later in AD (Jack et al., 2010). These markers of neurodegeneration, rather than those of Aβ or Tau proteinopathy, appear directly related to cognitive symptoms (Jack et al., 2010). Neurobiology of AD relates to axonal and neuronal degeneration followed by fibrillar lesions triggered by amyloid precursor protein (APP)-initiated death-receptor mechanism and activation of tau (Holtzman et al., 2011; Nikolaev et al., 2009). Initial axonal degeneration may lead to grey matter tissue changes and finally to neuronal loss or atrophy resulting in cognitive and functional impairment. Since diffusion MRI uses water molecules as an endogenous tracer to probe tissue microstructure or properties (Beaulieu, 2002), it can detect subtle changes in microstructure tissue properties in AD. Previous studies have shown that decreased white matter integrity is associated with AD (Acosta-Cabronero et al., 2010; Douaud et al., 2011; Zhang et al., 2009).

A potentially powerful application of diffusion MRI to AD research is assessing axonal white matter tracts using tractography. Tractography is a computational reconstruction of white matter tracts using biophysical modeling of fiber orientations (Johansen-Berg and Behrens, 2006; Seehaus et al., 2013). Recent advances in computational methods have enabled more rigorous estimation of white matter tracts (Azadbakht et al., 2015; Ciccarelli et al., 2008; Shi and Toga, 2017; Sporns, 2011). In AD, human imaging of APP and tau shows widespread topography. Given this, when tractography is applied at the connectome level, this structural connectome data could be useful for assessing axonal or white matter abnormalities across the entire connectome. A few studies using tractography at the connectome level have noted abnormal topological organization of structural connectome in AD (Dai and He, 2014; Lo et al., 2010). However, it remains untested whether and to what extent the structural connectome carries additional information that structural MRI and morphometry analysis do not present.

In this study, we addressed this issue using rigorous, data-driven machine learning in two independent datasets of moderate sample sizes (211 elders for the first dataset [Korean National Health Insurance Service Ilsan Hospital, South Korea] and 179 elders for the second, generalizability dataset [ADNI-2]). In both data, using multi-modal brain MRI (structural and diffusion MRI), we performed high-throughput brain phenotyping, including automated morphometry and white matter structural connectomics (probabilistic tractography) to generate large-scale multi-modal, multi-parametric imaging-derived phenotypes used as features in machine learning. A well-established, rigorous analysis pipeline was applied to diffusion MRI to estimate robust, individualized structure connectomes. We compared data-driven machine

learning classifiers trained on the individualized brain connectome and morphometric estimates with benchmark models (white matter hyperintensity) for the first Korean data and CSF biomarkers for the second reproducibility ADNI-2 data) using existing metrics.

MATERIALS AND METHODS

Participants. For the first study, we used data from 211 seniors who visited the dementia clinic at National Health Insurance Service Ilsan Hospital (NHIS-IH), Goyang, South Korea from 2010 to 2015. This sample is a randomly selected subset of the Ilsan Dementia Cohort, a retrospective clinical cohort. Neurologists made a diagnosis based on possible AD and Peterson's MCI criteria (Petersen, 2004), clinical history, a full battery of neuropsychological evaluations (Seoul neuropsychological screening battery) and MMSE (Mini-Mental State Examination). Those with vascular changes were not excluded from the study as long as they had a diagnosis of AD or MCI based on MMSE, CDR, and the neuropsychological evaluations. Those with AD as a primary diagnosis and with small vessel disease were noted as "AD with small vessel disease". Participants included 110 with the diagnosis of Alzheimer's disease (AD; median age=82; interquartile intervals (Q3-Q1)=85-77), 64 with mild cognitive impairment (MCI; median age=73; Q3-Q1=77-66), and 37 subjective memory complaints (SMC; median age=74; Q3-Q1=78-72) (**Table 1**). To test the generalizability of our approach, we also used ADNI-2 (Alzheimer's Disease Neuroimaging Initiative), where structural and diffusion MRI was collected. Demographical information is also provided in **Table 2**. The institutional review board of our hospital approved this study before implementation (IRB number:2017-04-24)

MRI acquisition. National Health Insurance Service Ilsan Hospital (NHIS-IH): We collected the following multimodal MRI from all participants: T1- MPRAGE: TE, 4.6 ms; matrix, 310 × 480× 480; voxel size, 0.5 × 0.5 × 0.5 mm. T2-FLAIR; matrix = 320 × 240 × 240; voxel size = 0.56 x 1.04 x 1.04. Diffusion MRI: matrix = 112 × 112 × 70; voxel size = 1.9 × 1.9 × 2.0 mm; the series included one image acquired without diffusion weighting and with diffusion weighting along 40 non-collinear directions (b = 600 s/m−2). ADNI-2: T1-weighted anatomical MRI and diffusion MRI. T1-MPRAGE: TE, min full echo; matrix, 208 × 240× 256; voxel size, 1 × 1 × 1 mm. Diffusion MRI: matrix = 256 × 256 × 46; voxel size = 1.36 × 1.36 × 2.7 mm; the series included 5 image acquired without diffusion weighting and with diffusion weighting along 41 non-collinear directions (b = 1000 s/m−2).

MRI Analysis-Structural MRI.

The high-throughput computational analysis was conducted. First, we estimated morphometric estimates using the Freesurfer image analysis pipeline (Fischl, 2012) (v6) from T1 and T2-FLAIR images. Morphometric measures (N=948 per subject) include volumes of the hippocampal subdivisions, and thickness, surface area, and volume of cortical/subcortical regions using two different atlases available in Freesurfer (Desikan-Killiany atlas and Destrieux atlas; https://surfer.nmr.mgh.harvard.edu/fswiki/CorticalParcellation). The technical details of these procedures are described in previous studies (Desikan et al., 2006; Destrieux et al., 2010; Fischl and Dale, 2000; Fischl et al., 1999). In brief, the image processing includes motion correction, removal of non-brain tissue, Talairach transformation, segmentation, intensity normalization, tessellation of the gray matter-white matter boundary, topology correction, and surface deformation. Deformation procedures use both intensity and continuity information to produce representations of cortical thickness. The maps produced are not restricted to the voxel resolution and are thus capable of detecting submillimeter differences between groups.

MRI Analysis-Diffusion MRI

We estimated structural connectome using the diffusion MRI analysis pipeline, MRtrix 3 (Tournier et al., 2004). The connectome measures (33,698 features per subject) include counts of streamlines, a surrogate measure of structural connectivity (Cha et al., 2015; Cha et al., 2017; Cha et al., 2016), and mean length of streamlines given any two brain regions based on multiple atlases. Diffusion-weighted magnetic resonance imaging (DWI) was preprocessed using the following pipeline in MRtrix 3. DWI was first denoised using a novel algorithm based on random matrix theory that permits data-driven, non-arbitrary threshold for Principal Component Analysis denoising; this method enhances the DWI quality for quantitative and statistical interpretation (Veraart et al., 2016). Denoised images then underwent eddy current and motion correction (Andersson and Sotiropoulos, 2016), brain extraction from three non-diffusion-weighted images (taking their median), and bias field correction using N4 algorithm (N4ITK), an improved N3 method, in Advanced Normalization Tools (ANTs)(Tustison et al., 2010). We then estimated fiber orientation distributions from each preprocessed image using 2nd-order integration over fiber orientation distributions (iFOD2). Based on the FODs, probabilistic tractography was performed using constrained spherical devolution (CSD). We used a target streamline count of 10 million across the whole brain. The tractograms were filtered using spherical-deconvolution informed filtering of tractograms (SIFT) with a target streamline count of 3 million. After a primary statistical analysis using these filtered tractograms, we tested whether the effects of interest were robust to the tractography and filtering parameters, such as the target streamline count for tractography, SIFT, or a ratio between them. This method permits mapping to streamline estimation back to individual's DWI and updating a reconstruction to improve model fit. This approach renders the streamline counts connecting two brain regions proportional to the total cross-sectional area of the white matter fibers connecting those regions, enhancing streamline counts as a biologically plausible quantity, representing "structural connectivity". This was done by repeating tractography and SIFT with a set of extreme parameters (100 million and 5 million target streamlines, respectively) with a filtering factor of 20 (100/5). Finally, from the filtered tractograms, we generated a connectivity matrix in each participant using two different atlases available in Freesurfer (Desikan-Killiany atlas (Desikan et al., 2006) and Destrieux atlas (Destrieux et al., 2010). We used streamline counts as the primary connectivity metric in this study as in a recent human infant imaging study (van den Heuvel et al., 2015b), as well mean length as secondary measures. A prior macaque study suggests the validity of streamline counts as an indicator of fiber connection strength, with the number of streamlines significantly correlating with tract-tracing strength in the macaque brain (van den Heuvel et al., 2015a).

Machine Learning Classification

We built several machine learning models using the large-scale brain MRI-derived phenotypes to predict the diagnosis of AD and MCI, respectively. Machine learning models were trained and cross-validated within each dataset. We benchmarked three commonly used classifiers available at a python library for machine learning, scikit-learn (Abraham et al., 2014): random forest, logistic regression (LR) with L1 and L2 regularization, and support vector machine (SVM) with a linear kernel. As a common preprocessing step for machine learning estimators, we

standardized the imaging derived phenotypes by removing the median and scaling them according to the quantile range (i.e., between the 1st and the 3rd quartile); this method is known to be robust to outliers (RobustScaler in scikit-learn python library). Model training and validation were done using nested cross-validation to avoid overfitting due to bias to training data (Cawley and Talbot, 2010; Varoquaux et al., 2017). Nested cross-validation uses a series of train/validation/test set splits: In the inner loop, we trained the model and selected a set of hyperparameters using the training set, then optimized the model with validation set; In the outer loop, we estimated generalization error of the underlying model using test sets. For feature selection, we used the 'forests of randomized trees' method, an ensemble method to combine the predictions of base estimators built with a learning algorithm, and then tested whether additional PCA-based dimensionality reduction improved the model or not. For hyper-parameter optimization, we used the grid search method, varying C parameter for SVM and LR classifier, and varying the number of estimators and the minimum samples per leaf for random forest classifier. We used nested, k-fold, stratified cross-validation with ten iterations. To avoid information leakage during cross-validation, our nested cross-validation scheme used a series of train/validation/test set splits. First, in the inner loop, feature selection was performed, and the model was trained in a train set, and the model performance was maximized via hyper-parameter optimization in a validation set. Secondly, in the outer loop, the model performance was evaluated in a test set, and generalization error was estimated by averaging test set scores across cross-validation splits.
To measure model performance, we used accuracy, sensitivity, specificity, and Area Under the Curve in receiver operating characteristic (AUC ROC). In diagnostic classification, we tested six different binary classifications, AD (coded as 1) vs. SMC (coded as 0), AD vs. MCI, MCI vs. SMC, AD only vs. AD with small vessel diseases, AD only vs. MCI, AD only vs. SMC.

Benchmark models

We used existing biomarkers as benchmark models. First, white matter hyperintensity in the Korean NHIS-IH study, and CSF biomarkers in the ADNI-2 study. White matter hyperintensity measures were estimated from T2-weighted FLAIR images using Wisconsin White Matter Hyperintensities Segmentation Toolbox (Ithapu et al., 2014). This method uses supervised machine learning methods to segment hyperintense regions and generates normalized effective white matter hyperintensity volume. Second, in ADNI-2 data, we used CSF biomarkers (phosphorylated tau, total tau, AB, ratio of phosphorylated tau/AB, ratio of total tau/AB), whose utility as biomarkers for diagnosis of AD (Olsson et al., 2016), MCI, and progression to AD from MCI (Hansson et al., 2006) has been studied. Furthermore, CSF biomarkers are reported to precede symptom onset of MCI (Moghekar et al., 2013).

RESULTS

Classification of AD and MCI

In the first study, we tested machine learning classification using white matter structural connectomes and morphometric estimates in 211 elders at the dementia clinic at the Korean National Health Insurance Service Ilsan Hospital. Age and sex alone showed moderate accuracies: AD/SMC: accuracy = 0.77; MCI/SMC: accuracy = 0.63; AD/MCI: accuracy = 0.72. White matter hyperintensity (WMH) served as a benchmark model, for it has been widely tested in the literature.

In classification of AD vs. SMC, optimal classification performance was shown in "morphometry+connectome" model (accuracy = 0.97, 95% CI=0.95-0.98) and "connectome" model (accuracy = 0.97, 95% CI=0.96-0.98) (**Table 3; Figure 1A**). These two models outperformed "morphometry" (accuracy = 0.87, 95% CI=0.85-0.88) and WMH benchmark models (accuracy = 0.73, 95% CI=0.71-0.75). In classification of MCI vs. SMC, similar classification performance was observed in "morphometry+connectome" (accuracy = 0.82, 95% CI=0.80-0.85) and "connectome" models (accuracy = 0.83, 95% CI=0.81-0.85), compared with lower performance of "morphometry" (accuracy = 0.59, 95% CI=0.57-0.60) and the WMH benchmark models (accuracy = 0.57, 95% CI=0.54-0.60). In classification of AD vs. MCI, "morphometry+connectome" models showed a best accuracy (accuracy=0.97, 95% CI=0.96-0.98), followed by "connectome" model (accuracy = 0.96, 95% CI=0.95-0.97), "morphometry" model (accuracy = 0.83, 95% CI=0.80-0.86), and the WMH benchmark models (accuracy = 0.66, 95% CI=0.64-0.69). Throughput all classifications, connectomes and morphometry showed greater diagnostic accuracies compared with the WMH benchmark.

Testing generalizability

We next tested the generalizability of the same multimodal brain imaging-based machine learning using ADNI-2 data. We included participants in ADNI-2 data whose structural and diffusion MRI (baseline) were both collected . To compare the performance of our classifiers, we used the invasive CSF biomarkers (p-tau, t-tau, Aβ42, p-tau/ Aβ42, t-tau/ Aβ42) as a benchmark model. In the classification of AD vs. HC, all the MRI-based models showed similarly optimal performance around 0.88 accuracy (**Table 4; Figure 1B**), outperforming the CSF benchmark model (accuracy = 0.75, 95% CI=0.73-0.77). In classification MCI vs. HC, all the MRI-based models showed similar performance with accuracies ranging from 0.64-0.67, outperforming the CSF benchmark (accuracy = 0.62, 95% CI=0.59-0.65). In classification AD vs. MCI, all the MRI-based models showed similar performance with accuracy ranging from 0.66-0.71, outperforming the CSF benchmark (accuracy = 0.54, 95% CI=0.52-0.57) which is barely above chance. This generalizability data showed, firstly, morphometry and connectome estimates showed equally good performance consistently exceeding the invasive CSF biomarkers in classifying AD/MCI/HC; secondly, unlike the NHIS-IH results, synergistic effects of combined morphometry and connectomes were not observed using our machine learning framework.

Testing utility for prognosis

Of the ADNI-2 data, we further tested the utility of our approach in predicting the disease trajectory. Data from 60 elders were used, whose baseline diagnosis was MCI and who were followed for at least two years. Machine learning models trained on the same five CSF benchmarks were used as a benchmark. In predicting progression from MCI to AD, “morphometry” model showed a highest accuracy (accuracy = 0.69, 95% CI=0.65-0.73) among MRI-based models, similar to the CSF benchmark model (accuracy = 0.70, 95% CI=0.66-0.75). (**Table 5, Figure 2**). “Connectome” model showed a lower, but statistically meaningful accuracy (accuracy = 0.57, 95% CI=0.53-0.61). Combining the two modalities of morphometry and connectomes (“morphometr+connectome”) did not improve the model accuracy (accuracy = 0.59, 95% CI=0.56-0.62), compared with “morphometry” model.

DISCUSSION

In this study, we used large-scale MRI-derived brain phenotypes (morphometry and white matter structural connectomes) with machine learning techniques to improve AD and MCI diagnosis in two independent Alzheimer's disease datasets compared with the benchmark models. We also predicted disease progression to AD from MCI. For high-throughput imaging analysis, we used a well-established automated pipeline for morphometry (Freesurfer 6.0) and a pipeline to estimate rigorously individualized white matter structural connectomes with up-to-date pre-/post-processing algorithms. Firstly, the models trained on morphometry and connectomes showed the best accuracy in classifying AD, MCI, and SMC or HC in the single-site data (ranging from 90% to 99% in AUC ROC; NHIS-IH, South Korea) as well as the multi-site (ranging from 70% to 97% in AUC ROC; ADNI-2, USA) "reproducibility" data. In most of the cases, the models outperformed the benchmark models significantly (e.g., white matter hyperintensity or CSF biomarkers) and demographic model (including age, sex, and education). Second, the model trained on connectome or morphometric estimates showed moderate accuracies (ranging from 57% to 79%; AUC) in predicting progression to AD in 60 elders with MCI in ADNI-2 data. These results show potential utility of white matter structural connectomes in addition to widely use morphometry as a proxy measure of neurodegeneration in AD pathology.

In two different datasets, our white matter structural connectome estimates showed promising results in classifying baseline diagnosis or predicting prognosis (MCI to AD progression). In the first study with NHIS-IH data, the models trained on connectome estimates outperformed morphometric estimates alone, as well as the existing benchmark metrics of white matter hyperintensity. This pattern was not so apparent in ADNI-2 data, where morphometric and connectome models show the similar model performance, and the combined models did not show a synergistic effect. Overall, this indicates the white matter structural connectomes provide essential information about brain pathophysiology of AD along with structural MRI-derived morphometry.

On the other hand, in the second study with ADNI-2 data, in predicting disease progression from MCI to AD, we found the morphometry model outperformed both connectome and combined models. This may suggest that grey matter morphometry provides more useful information in predicting the AD trajectory than the connectome measures. However, given the smaller sample size, machine learning training and feature selection may be suboptimal for the connectome model with a significantly larger number of features than the morphometry model; this might work favorably to the morphometry model. Furthermore, while morphometry and connectome models respectively showed statistically meaningful (above chance) predictions, when combined, there was little improvement in model performance. This indicates more rigorous methods to combine models trained across multimodal brain imaging-derived phenotypes, such as ensemble methods (Zhang et al., 2011), may be required.

A novel aspect of this study is to assess the utility of the dMRI-based white matter structural connectomes in predictive modeling of AD in a sufficiently large sample (n=211) and to validate

it in an independent study (n=179). Compared with other imaging modalities diffusion MRI and connectomics approach have been relatively understudied. Some prior studies have provided novel insights about the associations of structural connectomes with potential AD pathology (e.g., topological disturbance based on graph theory) (Pereira et al., 2017) and with healthy aging (Perry et al., 2015). Another line of literature shows the promising utility of connectomics estimates in predicting risk for AD, but with a caveat of limited samples sizes (n<30 (Wee et al., 2012; Zhu et al., 2014)). In line with this, this study presents validated results using larger samples from two independent studies. Also, the use of high-performance computing permitted to estimate individualized structural connectomes in a rigorous way such that it was not bound to limits of computational resources (e.g., restriction of tractography parameters). Therefore, this data may support a potential utility of the high-throughput brain phenotyping based on structure and diffusion MRI towards an accurate predictive model of risk for AD.

Across the two studies, we found the classification performance of the combined connectome and morphometry models were different. Though they generally outperformed the benchmark models in predicting AD or MCI, the combined model showed ~10% decrease in model performance in ADNI data compared with NHIS-IH data. Several points should be noted. First, since ADNI MRI came from multiple sites, dMRI are most likely to be affected by site effects. Similar to any MRI, dMRI signal characteristics change across sites; this may persist even when acquisition protocols are matched (cf. given there are no dynamic phantoms available for dMRI, there is hardly any objective ways to assess harmonization of dMRI acquisition). One way to mitigate the site effects (also as known as batch effects in biomedicine) is a statistical adjustment; a recent study suggests a Bayesian method for post-acquisition harmonization of dMRI (Fortin et al., 2017). However, despite its potential utility, more thorough assessment of such methods may be needed to determine its strengths and limitations (e.g., whether minimum batch sizes required, robustness to distinguish neurobiological signals vs. site effects). In this study, we did not use such a method, because no such methods were available for raw diffusion weighted images or FOD (fiber orientation distribution) maps (so that the tractography and connectome estimation may benefit from harmonized data) used for estimation of the structural connectomes. Secondly, a socio-demographic difference between the NHIS-IH and the ADNI studies may be related to the difference in model performance. Indeed, ADNI participants have a higher education than NHIS-IH participants; this is probably because NHIS-IH participants were affected by the Korean war in their 10-20's. Accordingly, we speculate that ADNI cohort might have a more substantial cognitive reserve than NHIS-IH, in which case the white matter structural connectomes in the ADNI cohort might be better protected even in the case of AD, compared with the NHIS-IH cohort.

Regarding the differences in model performance across the datasets, one innovative way is to use machine learning is for domain-invariant feature learning. For example, a recent seminal study (Ghafoorian et al., 2017) of white matter hyperintensity segmentation in the brain shows a successful application of domain adaption: a convolutional neural network was trained on data from a single domain (i.e., from a single scanner with a single acquisition protocol), which was then retrained (transfer learning) on the same task with independent brain images from different domains (i.e., different acquisition protocols and image dimension from the same scanner).

Given the revolutionizing success of deep learning and the large volume of multi-center brain imaging data available for AD research, deep neural network-based domain adaptation might be a promising area of research towards a generalizable predictive model in AD.

Our samples from NHIS-IH data represents a retrospective clinical cohort. Compared to some of the publicly available imaging databases for AD research, our clinical data may present a greater heterogeneity in aging samples; therefore, these samples may well represent a true clinical population. Given this, it is interesting such an accurate classification for both AD and MCI. However, in classifying AD/SMC, given the significant difference in age (i.e., higher in the AD group), it is likely that a greater aging effect embedded on morphometric and connectome estimates contributed to the perfect classification. Owing to the lack of age-matched healthy controls, in this study we could not disentangle the impact of normal and pathological (i.e., AD) aging on brain phenotypes. This remains to be tested in future research.

On the other hand, the accurate classification of MCI/SMC (age-matched) is notable. Our analysis showed the most significant contributor to this model was structural connectome. The connectome model classified MCI/SMC as equally accurate as the combined model, whereas the morphometry model did not classify accurately. This pattern was particularly more pronounced in MCI/SMC classification than in AD/SMC classification. This may reflect different time ordering of white matter integrity (or connectivity) and grey matter atrophy. Literature shows the capability of diffusion MRI-derived measures to detect subtle changes in tissue properties or microstructure, whereas structural MRI is typically used to estimate macroscopic properties, namely volume.

Our results show white matter structural connectomes and morphometry contains salient information perhaps complementary to each other and important for prediction of clinical outcomes in AD. This study may thus lend a support to the future research using the white matter connectomics in AD research. Future research should test rigorous analytics, such as deep neural networks, that could optimally use the large-scale connectomic and morphometric information. In the future the combination of multimodal brain MRI and rigorous analytics might permit to detect subtle brain abnormalities before the onset of cognitive symptoms or severe neurodegeneration (e.g., hippocampal volume loss).

Acknowledgments

This work used the Extreme Science and Engineering Discovery Environment Stampede 2 at the Texas Advanced Computing Center (TG-IBN170015: Cha) and Argonne National Laboratory Leadership Computing Facility (PI, Cha). This study was supported by NIMH K01 MH109836 (Cha), Brain and Behavior Research Foundation NARSAD Young Investigator award (Cha), Korean Scientists and Engineers Association Young Investigator Grant (Cha), National health insurance Ilsan hospital research fund.

Tables

Table 1. NHIS-IH Participant Demographics

	AD (N= 110)	MCI (N=62)	SMC (N=36)	Test Statistics	P value
Age,Mean (SD)	79.95 (6.61)	71.42 (8.62)	72.25 (6.99)	F = 32.72	P < 0.001
Sex					
Female	74	38	32	χ^2 =8.56	P = 0.014
Male	36	24	4		
Education	6.7 (5.2)	9.8 (4.6)	7.6 (4.9)	F = 6.541	P = 0.011
MMSE	18.1 (0.53)	25.1 (0.36)	26.3 (0.37)	F = 151.9	P < 0.001
CDR	1.03 (0.57)	0.54 (0.13)	0.50 (0.11)	F = 79.38	P < 0.001

NHIS-IH, National Health Insurance Service Ilsan Hospital; **SD**, standard deviation; **MMSE,** Mini Mental State Examination; **CDR**, the clinical Dementia Rating.

Table 2. ADNI-2 Participant Demographics

	AD (N=48)	MCI (N=60)	HC (N= 71)	Test Statistics	P value
Age,Mean (SD)	74.96 (8.59)	72.57 (6.62)	72.55 (5.66)	F = 3.11	P =0.08
Sex					
Female	20	20	43	χ^2 =10.28	P =0.006
Male	28	40	28		
Education	15.31 (2.87)	16.08 (2.68)	16.28 (2.72)	F = 6.541	P = 0.07
CDR	0.82 (0.24)	0.50 (0.00)	0	F=663.1	P < 0.001

ADNI-2, Alzheimer's disease neuroimaging Initiative; **SD**, standard deviation; **CDR**, the clinical Dementia Rating.

Table 3. Performances of Machine Learning Classifier using NHIS-IH Structural Connectomes, Morphometric Brain Features, and white matter hyperintensity.

		AD vs SMC	MCI vs SMC	AD vs MCI
Morphometry + Connectome	**AUC**	**0.99**(0.99-1.00) ♠	**0.90**(0.87-0.92) ♠	**0.99**(0.98-1.00)♠
	Accuracy	**0.97**(0.95-0.98)	**0.82**(0.80-0.85)	**0.97**(0.96-0.98)
	Sensitivity	**0.94**(0.92-0.97)	**0.76**(0.72-0.80)	**0.96**(0.94-0.97)
	Specificity	**0.98**(0.98-0.99)	**0.87**(0.85-0.90)	**0.98**(0.97-0.99)
Connectome	**AUC**	**0.99**(0.99-1.00) ♠	**0.90**(0.88-0.92) ♠	**0.99**(0.99-1.00) ♠
	Accuracy	**0.97**(0.96-0.98)	**0.83**(0.81-0.85)	**0.96**(0.95-0.97)
	Sensitivity	**0.94**(0.93-0.97)	**0.77**(0.74-0.81)	**0.94**(0.91-0.96)
	Specificity	**0.98**(0.97-0.99)	**0.88**(0.85-0.91)	**0.98**(0.97-0.99)
Morphometry	**AUC**	**0.88**(0.86-0.90)	**0.48**(0.45-0.50)	**0.85**(0.82-0.88)
	Accuracy	**0.87**(0.85-0.88)	**0.59**(0.57-0.60)	**0.83**(0.80-0.86)
	Sensitivity	**0.84**(0.79-0.88)	**0.33**♣	**0.80**(0.75-0.86)
	Specificity	**0.88**(0.86-0.89)	**0.63**(0.62-0.64)	**0.85**(0.82-0.86)
Benchmark (White Matter Hyperintensity)	**AUC**	**0.67**(0.64-0.70)	**0.45**(0.42-0.49)	**0.61**(0.57-0.64)
	Accuracy	**0.73**(0.71-0.75)	**0.57**(0.54-0.60)	**0.66**(0.64-0.69)
	Sensitivity	**0.38**♣	**0.26**♣	**0.44**♣
	Specificity	**0.78**(0.76-0.79)	**0.61**(0.60-0.63)	**0.72**(0.70-0.73)

NHIS-IH, National Health Insurance Service Ilsan Hospital; **SMC**, subjective memory complaints**; MCI,** mild cognitive impairment; **AD,** Alzheimer's disease.*All results show mean and standard deviation as **mean** and **95% confidence interval** in this table. ♠ indicates the best models for this classification; ♣ indicates the confidence interval is not available. For all three classifications, random forest performed as the best classifier, therefore, we only put random forest classifier performance results into this table.

Table 4. Performances of Machine Learning Classifiers using ADNI-2 Structural Connectomes and Morphometric Brain Features.

		AD vs HC	MCI vs HC	AD vs MCI
Morphometry + Connectome	**AUC**	**0.96**(0.94-0.97)	**0.70**(0.67-0.73)	**0.75**(0.72-0.78)
	Accuracy	**0.88**(0.86-0.90)	**0.64**(0.61-0.66)	**0.66**(0.64-0.69)
	Sensitivity	**0.89**(0.87-0.90)	**0.64**(0.62-0.66)	**0.69**(0.68-0.72)
	Specificity	**0.91**(0.88-0.93)	**0.66**(0.62-0.70)	**0.65**♣
Connectome	**AUC**	**0.95**(0.94-0.96)	**0.72**(0.69-0.75)♠	**0.75**(0.73-0.78)
	Accuracy	**0.89**(0.87-0.90)	**0.64**(0.62-0.67)	**0.67**(0.65-0.70)
	Sensitivity	**0.88**(0.87-0.90)	**0.65**(0.63-0.67)	**0.71**(0.68-0.74)
	Specificity	**0.92**(0.89-0.94)	**0.66**♣	**0.67**(0.63-0.71)
Morphometry	**AUC**	**0.97**(0.96-0.98)♠	**0.71**(0.67-0.74)	**0.79**(0.76-0.81)♠
	Accuracy	**0.89**(0.87-0.91)	**0.67**(0.64-0.69)	**0.71**(0.69-0.73)
	Sensitivity	**0.88**(0.86-0.90)	**0.67**(0.65-0.69)	**0.73**(0.70-0.75)
	Specificity	**0.92**(0.89-0.95)	**0.69**♣	**0.70**(0.66-0.73)
Benchmark (CSF Biomarkers)	**AUC**	**0.79**(0.77-0.82)	**0.65**(0.62-0.68)	**0.56**(0.53-0.59)
	Accuracy	**0.75**(0.73-0.77)	**0.62**(0.59-0.65)	**0.54**(0.52-0.57)
	Sensitivity	**0.76**(0.74-0.78)	**0.63**(0.60-0.66)	**0.59(0.56-0.61)**
	Specificity	**0.76**(0.72-0.80)	**0.61**(0.57-0.65)	**0.49**(0.46-0.52)

ADNI-2, Alzheimer's Disease Neuroimaging Initiative **2; MCI,** mild cognitive impairment; **AD,** Alzheimer's disease**; HC**, healthy control. *All results show Mean and standard deviation as **mean** and **95% confidence interval** in this table. ♠ indicates the best models for this classification. For all three classifications, random forest performed as the best classifier, therefore, we only put random forest classifier performance results into this table. ♣ indicates the confidence interval is not available.

Table 5. Prognostic Performances of Machine Learning Classifier using ADNI-2

MCI-AD vs. Stable MCI	
Morphometry	
Accuracy	0.69 (0.65-0.73)*
Sensitivity	0.79 (0.74-0.83)
Specificity	0.69 (0.64-0.74)
AUC	0.79 (0.74-0.84)
Connectomes	
Accuracy	0.57 (0.53-0.61)
Sensitivity	0.64 (0.58-0.69)
Specificity	0.53 (0.47-0.59)
AUC	0.62 (0.56-0.68)
Morphometry + Connectome	
Accuracy	0.59 (0.56-0.62)
Sensitivity	0.60 (0.56-0.63)
Specificity	0.68 (0.56-0.79)
AUC	0.65 (0.59-0.71)
Benchmark: CSF biomarkers	
Accuracy	0.70 (0.66-0.75)
Sensitivity	0.76 (0.72-0.81)
Specificity	0.71 (0.64-0.78)
AUC	0.76 (0.70-0.81)

ADNI-2, Alzheimer's Disease Neuroimaging Initiative **2; MCI,** mild cognitive impairment; **AD,** Alzheimer's disease**; LR**, logistic regression; **PCA**, principal component analysis; **CV**, cross-validation. *All results show Mean and standard deviation as **mean** and **95% confidence interval** in this table.

Figures

Figure 1. Classification of baseline diagnosis using connectomes and morphometric estimates. (*A*), classification performances in the first study (Korean National Health Insurance Ilsan Hospital data).It showed higher diagnostic accuracy (area under the curve of the receiver-operator characteristics or AUC ROC) of the machine learning model trained on combined connectome and morphometric estimates consistently, compared with the benchmark model trained on white matter hyperintensity. Out of three machine learning algorithms (random forest, support vector machine, and logistic regression), best models were shown. (***B***), classification performances in the second study (ADNI-2). It showed reproducible results of diagnostic accuracy of connectomes and morphometry. The combined models show better performance in predicting AD from healthy controls and AD from MCI, and similar in predicting MCI from HC. Best models were shown. Compared with the first study, the reproducibility data shows less diagnostic accuracy presumably due to multiple sites and stricter inclusion and exclusion criteria in ADNI. **WMH**, white matter hyperintensity; **Demo**, demographics including sex, age, and education.

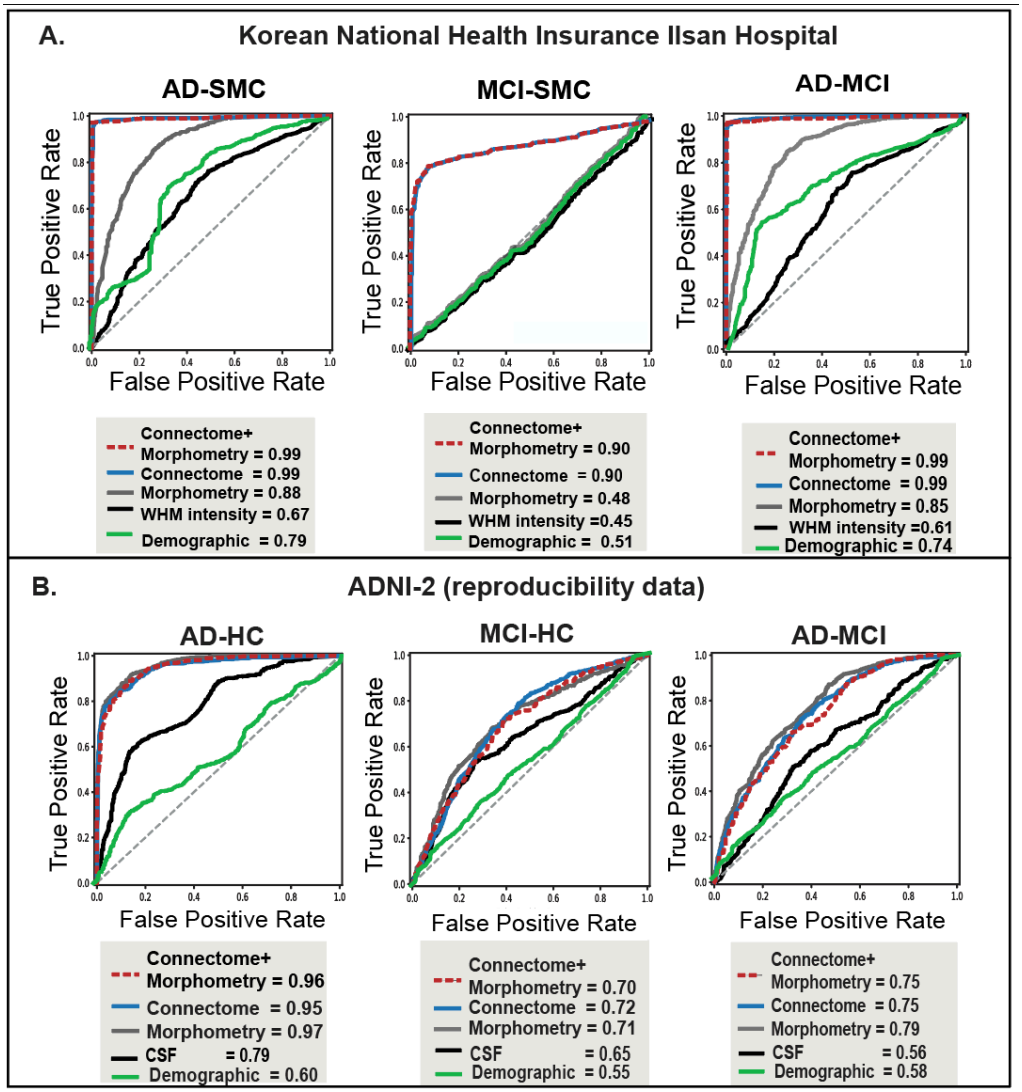
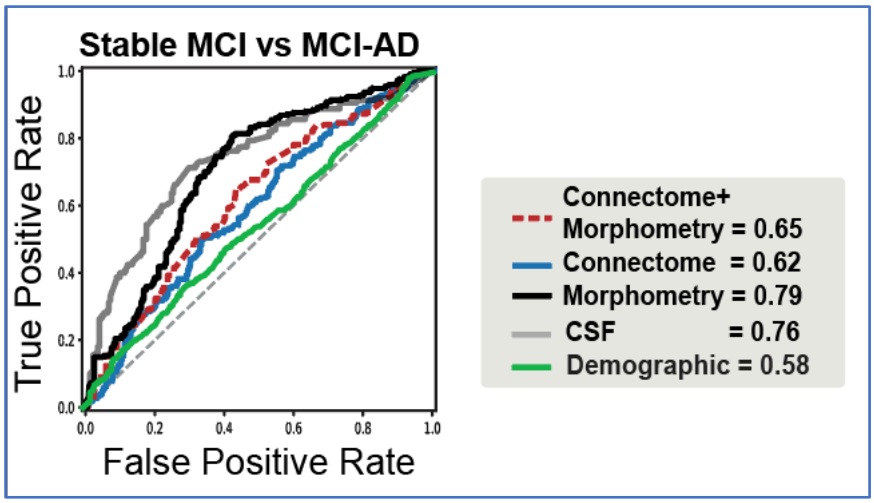


Figure 2. Prediction of progression to AD from MCI using connectomes and morphometric estimates. Using ADNI-2 data that has follow-up data after baseline MRI scan, machine learning models were tested using connectome and morphometry estimates to predict MRI-to-AD progression in 50 elders with MCI (mean follow-up years in stable MCI, 3.76 ± 0.98; range, 2.18-5.32). Morphometry model showed similar performance to CSF benchmark model. Both the combined model and connectome model showed lower but meaningful accuracy.



References

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., Varoquaux, G., 2014. Machine learning for neuroimaging with scikit-learn. Front Neuroinform 8, 14.

Acosta-Cabronero, J., Williams, G.B., Pengas, G., Nestor, P.J., 2010. Absolute diffusivities define the landscape of white matter degeneration in Alzheimer's disease. Brain 133, 529-539.

Andersson, J.L., Sotiropoulos, S.N., 2016. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. Neuroimage 125, 1063-1078.

Azadbakht, H., Parkes, L.M., Haroon, H.A., Augath, M., Logothetis, N.K., de Crespigny, A., D'Arceuil, H.E., Parker, G.J., 2015. Validation of High-Resolution Tractography Against In Vivo Tracing in the Macaque Visual Cortex. Cereb Cortex 25, 4299-4309.

Beaulieu, C., 2002. The basis of anisotropic water diffusion in the nervous system - a technical review. NMR Biomed 15, 435-455.

Cawley, G.C., Talbot, N.L.C., 2010. On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. Journal of Machine Learning Research 11, 2079-2107.

Cha, J., Fekete, T., Siciliano, F., Biezonski, D., Greenhill, L., Pliszka, S.R., Blader, J.C., Roy, A.K., Leibenluft, E., Posner, J., 2015. Neural Correlates of Aggression in Medication-Naive Children with ADHD: Multivariate Analysis of Morphometry and Tractography. Neuropsychopharmacology 40, 1717-1725.

Cha, J., Guffanti, G., Gingrich, J., Talati, A., Wickramaratne, P., Weissman, M., Posner, J., 2017. Effects of Serotonin Transporter Gene Variation on Impulsivity Mediated by Default Mode Network: A Family Study of Depression. Cereb Cortex, 1-11.

Cha, J., Ide, J.S., Bowman, F.D., Simpson, H.B., Posner, J., Steinglass, J.E., 2016. Abnormal reward circuitry in anorexia nervosa: A longitudinal, multimodal MRI study. Hum Brain Mapp 37, 3835-3846.

Ciccarelli, O., Catani, M., Johansen-Berg, H., Clark, C., Thompson, A., 2008. Diffusion-based tractography in neurological disorders: concepts, applications, and future developments. Lancet Neurol 7, 715-727.

Dai, Z., He, Y., 2014. Disrupted structural and functional brain connectomes in mild cognitive impairment and Alzheimer's disease. Neurosci Bull 30, 217-232.

Desikan, R.S., Segonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., 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.

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

Douaud, G., Jbabdi, S., Behrens, T.E., Menke, R.A., Gass, A., Monsch, A.U., Rao, A., Whitcher, B., Kindlmann, G., Matthews, P.M., Smith, S., 2011. DTI measures in crossing-fibre areas: increased diffusion anisotropy reveals early white matter alteration in MCI and mild Alzheimer's disease. Neuroimage 55, 880-890.

Fischl, B., 2012. FreeSurfer. Neuroimage 62, 774-781.

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

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.

Fortin, J.P., Parker, D., Tunc, B., Watanabe, T., Elliott, M.A., Ruparel, K., Roalf, D.R., Satterthwaite, T.D., Gur, R.C., Gur, R.E., Schultz, R.T., Verma, R., Shinohara, R.T., 2017. Harmonization of multi-site diffusion tensor imaging data. Neuroimage 161, 149-170.

Ghafoorian, M., Mehrtash, A., Kapur, T., Karssemeijer, N., Marchiori, E., Pesteie, M., Guttmann, C.R., de Leeuw, F.-E., Tempany, C.M., van Ginneken, B., 2017. Transfer learning for domain adaptation in mri: Application in brain lesion segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 516-524.

Hansson, O., Zetterberg, H., Buchhave, P., Londos, E., Blennow, K., Minthon, L., 2006. Association between CSF biomarkers and incipient Alzheimer's disease in patients with mild cognitive impairment: a follow-up study. Lancet Neurol 5, 228-234.

Holtzman, D.M., Morris, J.C., Goate, A.M., 2011. Alzheimer's disease: the challenge of the second century. Sci Transl Med 3, 77sr71.

Ithapu, V., Singh, V., Lindner, C., Austin, B.P., Hinrichs, C., Carlsson, C.M., Bendlin, B.B., Johnson, S.C., 2014. Extracting and summarizing white matter hyperintensities using supervised segmentation methods in Alzheimer's disease risk and aging studies. Hum Brain Mapp 35, 4219-4235.

Jack, C.R., Jr., Knopman, D.S., Jagust, W.J., Shaw, L.M., Aisen, P.S., Weiner, M.W., Petersen, R.C., Trojanowski, J.Q., 2010. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol 9, 119-128.

Johansen-Berg, H., Behrens, T.E., 2006. Just pretty pictures? What diffusion tractography can add in clinical neuroscience. Curr Opin Neurol 19, 379-385.

Lo, C.Y., Wang, P.N., Chou, K.H., Wang, J., He, Y., Lin, C.P., 2010. Diffusion tensor tractography reveals abnormal topological organization in structural cortical networks in Alzheimer's disease. J Neurosci 30, 16876-16885.

Moghekar, A., Li, S., Lu, Y., Li, M., Wang, M.C., Albert, M., O'Brien, R., Team, B.R., 2013. CSF biomarker changes precede symptom onset of mild cognitive impairment. Neurology 81, 1753-1758.

Nikolaev, A., McLaughlin, T., O'Leary, D.D., Tessier-Lavigne, M., 2009. APP binds DR6 to trigger axon pruning and neuron death via distinct caspases. Nature 457, 981-989.

Olsson, B., Lautner, R., Andreasson, U., Ohrfelt, A., Portelius, E., Bjerke, M., Holtta, M., Rosen, C., Olsson, C., Strobel, G., Wu, E., Dakin, K., Petzold, M., Blennow, K., Zetterberg, H., 2016. CSF and blood biomarkers for the diagnosis of Alzheimer's disease: a systematic review and meta-analysis. Lancet Neurol 15, 673-684.

Pereira, J.B., van Westen, D., Stomrud, E., Strandberg, T.O., Volpe, G., Westman, E., Hansson, O., 2017. Abnormal Structural Brain Connectome in Individuals with Preclinical Alzheimer's Disease. Cereb Cortex, 1-12.

Perry, A., Wen, W., Lord, A., Thalamuthu, A., Roberts, G., Mitchell, P.B., Sachdev, P.S., Breakspear, M., 2015. The organisation of the elderly connectome. Neuroimage 114, 414-426.

Petersen, R.C., 2004. Mild cognitive impairment as a diagnostic entity. Journal of internal medicine 256, 183-194.

Seehaus, A.K., Roebroeck, A., Chiry, O., Kim, D.S., Ronen, I., Bratzke, H., Goebel, R., Galuske, R.A., 2013. Histological validation of DW-MRI tractography in human postmortem tissue. Cereb Cortex 23, 442-450.

Shi, Y., Toga, A.W., 2017. Connectome imaging for mapping human brain pathways. Mol Psychiatry 22, 1230-1240.

Sporns, O., 2011. The human connectome: a complex network. Ann N Y Acad Sci 1224, 109-125.

Teipel, S., Drzezga, A., Grothe, M.J., Barthel, H., Chetelat, G., Schuff, N., Skudlarski, P., Cavedo, E., Frisoni, G.B., Hoffmann, W., Thyrian, J.R., Fox, C., Minoshima, S., Sabri, O., Fellgiebel, A., 2015. Multimodal imaging in Alzheimer's disease: validity and usefulness for early detection. Lancet Neurol 14, 1037-1053.

Tournier, J.D., Calamante, F., Gadian, D.G., Connelly, A., 2004. Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. Neuroimage 23, 1176-1185.

Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C., 2010. N4ITK: improved N3 bias correction. IEEE Trans Med Imaging 29, 1310-1320.

van den Heuvel, M.P., de Reus, M.A., Feldman Barrett, L., Scholtens, L.H., Coopmans, F.M., Schmidt, R., Preuss, T.M., Rilling, J.K., Li, L., 2015a. Comparison of diffusion tractography and tract-tracing measures of connectivity strength in rhesus macaque connectome. Hum Brain Mapp 36, 3064-3075.

van den Heuvel, M.P., Kersbergen, K.J., de Reus, M.A., Keunen, K., Kahn, R.S., Groenendaal, F., de Vries, L.S., Benders, M.J., 2015b. The Neonatal Connectome During Preterm Brain Development. Cereb Cortex 25, 3000-3013.

Varoquaux, G., Raamana, P.R., Engemann, D.A., Hoyos-Idrobo, A., Schwartz, Y., Thirion, B., 2017. Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. Neuroimage 145, 166-179.

Veraart, J., Novikov, D.S., Christiaens, D., Ades-Aron, B., Sijbers, J., Fieremans, E., 2016. Denoising of diffusion MRI using random matrix theory. Neuroimage 142, 394-406.

Wee, C.-Y., Yap, P.-T., Zhang, D., Denny, K., Browndyke, J.N., Potter, G.G., Welsh-Bohmer, K.A., Wang, L., Shen, D., 2012. Identification of MCI individuals using structural and functional connectivity networks. Neuroimage 59, 2045-2056.

Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D., Alzheimer's Disease Neuroimaging, I., 2011. Multimodal classification of Alzheimer's disease and mild cognitive impairment. Neuroimage 55, 856-867.

Zhang, Y., Schuff, N., Du, A.T., Rosen, H.J., Kramer, J.H., Gorno-Tempini, M.L., Miller, B.L., Weiner, M.W., 2009. White matter damage in frontotemporal dementia and Alzheimer's disease measured by diffusion MRI. Brain 132, 2579-2592.

Zhu, D., Li, K., Terry, D.P., Puente, A.N., Wang, L., Shen, D., Miller, L.S., Liu, T., 2014. Connectome-scale assessments of structural and functional connectivity in MCI. Human brain mapping 35, 2911-2923.

Supplemental Tables

Supplemental Table 1. Performances of Machine Learning Classifier using NHIS-IH Structural Connectomes, Morphometric Brain Features, and white matter hyperintensity.

		ADonly vs AD with small vessels	ADonly vs MCI	ADonly vs SMC
Morphometry + Connectome	**AUC**	**0.69**(0.65-0.72)	**0.82**(0.79-0.83) ♠	**0.80**(0.78-0.82) ♠
	Accuracy	**0.71**(0.69-0.73)	**0.75**(0.72-0.76)	**0.77**(0.76-0.79)
	Sensitivity	**0.41**♣	**0.77**(0.74-0.81)	**0.81**(0.77-0.84)
	Specificity	**0.73**(0.72-0.74)	**0.74**(0.72-0.76)	**0.77**(0.76-0.79)
Connectome	**AUC**	**0.64**(0.60-0.68)	**0.79**(0.77-0.81)	**0.78**(0.76-0.80)
	Accuracy	**0.70**(0.68-0.72)	**0.74**(0.72-0.76)	**0.76**(0.74-0.78)
	Sensitivity	**0.38**♣	**0.74**(0.71-0.78)	**0.79**(0.75-0.84)
	Specificity	**0.73**(0.72-0.74)	**0.75**(0.73-0.76)	**0.76**(0.74-0.77)
Morphometry	**AUC**	**0.69**(0.66-0.72)	**0.78**(0.76-0.81)	**0.80**(0.78-0.82) ♠
	Accuracy	**0.72**(0.70-0.73)	**0.70**(0.67-0.73)	**0.76**(0.74-0.78)
	Sensitivity	**0.52**♣	**0.65**(0.62-0.68)	**0.75**(0.70-0.80)
	Specificity	**0.75**(0.73-0.76)	**0.75**(0.72-0.77)	**0.77**(0.75-0.79)
Benchmark (White Matter Hyperintensity)	**AUC**	**0.96**(0.95-0.97) ♠	**0.50**♣	**0.62**(0.60-0.64)
	Accuracy	**0.90**(0.87-0.91)	**0.57**(0.56-0.58)	**0.65**(0.64-0.67)
	Sensitivity	**0.89**(0.85-0.94)	**0.73**♣	**0.54**♣
	Specificity	**0.90**(0.88-0.92)	**0.57**(0.56-0.58)	**0.69**(0.68-0.71)

NHIS-IH, National Health Insurance Service Ilsan Hospital; **SMC**, subjective memory complaints**; MCI,** mild cognitive impairment; **AD,** Alzheimer's disease.*All results show Mean and standard deviation as **mean** and **95% confidence interval** in this table. ♠ indicates the best models for this classification; ♣ indicates the confidence interval is not available. For all three classifications, random forest performed as the best classifier, therefore, we only put random forest classifier performance results into this table.

Notice: The best classifier for ADonly vs. AD with small vessels is an SVM classifier, and a logistic regression classifier performed best for ADonly vs. MCI, random forest classifier performed best for ADonly vs. SMC.