Published in final edited form as:

J Alzheimers Dis. 2018 ; 64(1): 149–169. doi:10.3233/JAD-171048.

The added value of diffusion-weighted MRI-derived structural connectome in evaluating mild cognitive impairment: a multi-cohort validation

Qi Wang¹, Lei Guo², Paul M. Thompson³, Clifford R. Jack Jr.⁴, Hiroko Dodge^{5,6}, Liang Zhan⁷, Jiayu Zhou¹, and Alzheimer's Disease Neuroimaging Initiative and National Alzheimer's Coordinating Center¹

¹Computer Science and Engineering, Michigan State University, East Lansing, MI

²Mathematics, Statistics & Computer Science Department, University of Wisconsin-Stout, Menomonie, WI

³Imaging Genetics Center, University of Southern California, Marina del Rey, CA

⁴Department of Radiology, Mayo Clinic, Rochester, MN

⁵Michigan Alzheimer's Disease Center and Department of Neurology, University of Michigan, Ann Arbor, MI

⁶Layton Aging and Alzheimer's Disease Center and Department of Neurology, Oregon Health & Science University, Portland, OR

⁷Computer Engineering Program, University of Wisconsin-Stout, Menomonie, WI

Abstract

T1-weighted MRI has been extensively used to extract imaging biomarkers and build classification models for differentiating AD from healthy controls, but only recently have brain connectome networks derived from diffusion-weighted MRI been used to model AD progression and various stages of disease such as mild cognitive impairment. Mild cognitive impairment, as a possible prodromal stage of AD, has gained intense interest recently, since it may be used to assess risk factors for AD. Little work has been done to combine information from both white matter and gray matter, and it is unknown how much classification power the diffusion-weighted MRI-derived structural connectome could provide beyond information available from T1-weighted MRI. In this paper, we focused on investigating whether diffusion-weighted MRI-derived structural connectome can improve differentiating healthy controls subjects from those with mild cognitive impairment. Specifically, we proposed a novel feature-ranking method to build classification

¹Data used in preparation of this article were obtained from the ADNI (adni.loni.usc.edu) and NACC (www.alz.washington.edu) respectively. As such, the investigators within the ADNI and NACC contributed to the design and implementation of respective databases 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 author: Dr. Jiayu Zhou, Department of Computer Science and Engineering, Michigan State University, 428 S Shaw Ln, East Lansing, MI 48824. jiayuz@msu.edu.

⁷Conflict of Interest/Disclosure Statement

The authors have no conflict of interest to report

models using the most highly ranked feature variables to classify mild cognitive impairment with healthy controls. We verified our method on two independent cohorts including the second stage of Alzheimer's Disease Neuroimaging Initiative (ADNI2) database and the National Alzheimer's Coordinating Center (NACC) database. Our results indicated that 1) diffusion-weighted MRI-derived structural connectome can complement T1-weighted MRI in the classification task; 2) the feature-rank method is effective because of the identified consistent T1-weighted MRI and network feature variables on ADNI2 and NACC. Furthermore, by comparing the top-ranked feature variables from ADNI2, NACC and combined dataset, we concluded that cross-validation using independent cohorts is necessary and highly recommended.

Keywords

Mild Cognitive Impairment; diffusion MRI; brain network; multiple cohorts; feature extraction; Alzheimer's disease

1. Introduction:

In Alzheimer's disease research, mild cognitive impairment (MCI) is commonly studied as a transitional stage between the cognitive decline expected with normal aging and the more serious decline of dementia. It can involve problems with memory, language, attention and judgment that are greater than normal age-related changes, but are not significant enough to interfere with daily activities. In [1], the authors analyzed how 5 most well-established AD biomarkers evolve in relation to each other and the onset and progression of clinical symptoms from healthy controls (HC), to MCI and dementia. Even so, it is not yet fully understood how the brain deteriorates from health to MCI and eventually to dementia, and understanding these mechanisms and processes is vital for evaluating preventions and treatments. In addition, more effective prognosis and evaluations for people with MCI could offer an enormous public health benefit.

Providing detailed information on the brain's gray matter, T1-weighted MRI (T1w MRI) has been widely used in discovering biomarkers related to AD. By averaging cortical feature variables in an AD population and in matched elderly controls, [2] identified striking profiles of gray matter loss, anatomical variation and some evidence of cerebral asymmetries in these deficits. Significant gray matter deficits were observed in a broad anatomical region encompassing bilateral temporal and parietal cortices in patients with AD [3]. Recently, machine learning techniques have been introduced to identify changing patterns of brain structures and improve prognostic accuracy. For example, [4-6] used Support Vector Machine (SVM) to differentiate AD from HC with discrete wavelet transform based features or displacement field based features, and achieved a promising classification accuracy. [7] adopted SVM and logistic regression and obtained reasonable accuracy in classifying individuals as HC versus MCI, and HC versus AD. In [8], the authors used voxel-based analyses and SVM-based pattern recognition to analyze 77 three-dimensional T1w MRI data sets from subjects with amnesic MCI and showed a significant cluster of gray matter density reduction, in the left hippocampal region. [9] formulated the disease progression prediction problem as a multi-task regression problem by considering the prediction at each

time point as a task. Their results showed that cortical thickness averages from the left middle temporal gyri, and the left and right entorhinal cortices, and a measure of white matter volume from the left hippocampus play significant roles in predicting decline in a key cognitive measure – the ADAS-Cog [10] - through the course of progression.

While T1w MRI has provided an effective way to capture structural information on the brain's gray matter, diffusion-weighted MRI (dMRI) is sensitive to microscopic properties of the brain's white matter that are not detectable with standard anatomical MRI, and is widely adopted in brain research. dMRI can provide maps of connections among different brain regions. This type of map is termed the brain's “structural connectome” or simply a brain network, which may include hundreds to thousands of nodes indicating brain regions-of-interest (ROIs) and the weighted connections (edges) connecting them. Brain structural network analysis has been widely used to study various brain diseases [11-13] including Alzheimer's [14-20]. By using a model of the brain's structural network, [19] reported that AD patients had significantly decreased nodal efficiency at the regional level, as well as weaker connections in multiple local cortical and subcortical regions, such as precuneus, temporal lobe, hippocampus, and thalamus. [16] tested how AD disrupts the ‘rich club’ effect. They found AD patients had a lower nodal degree in cortical regions implicated in the disease, as expected. The normalized rich club coefficient was higher in AD. In [14], the authors studied 34 participants, finding decreased fiber density and disrupted connectivity between the hippocampus and posterior cingulate cortex in early AD. The MCI group showed reduced fiber density from the posterior cingulate cortex and hippocampus to the rest of the brain.

The combination of different MRI modalities has been successfully applied for prognosis of some diseases [21-27], including MCI [28-31] [32-34]. In those studies, the authors used dMRI-derived features such as fractional anisotropy (FA) values, apparent diffusion coefficient (ADC) combined with T1w features such as cortical thickness, cortical volume, etc. to classifying MCI with HC. The results showed promising performance by combining different MRI modalities. However, few existing studies use dMRI-derived structural connectome and T1w information to differentiate MCI with HC. This encourages us to explore the effectiveness of combining structural connectome markers and T1w MRI markers. Also, most previous studies only validated their method on one dataset. We believe that cross-validation on multiple datasets can lead to insightful analysis. In this study, we tested the hypothesis that dMRI-derived structural connectome can supplement T1w MRI by boosting the classification performance in differentiating MCI subjects from HC subjects. Specifically, we extracted dMRI-derived network feature variables, i.e. the connections between 113 ROIs [35], and T1w MRI-derived feature variables including cortical volume and cortical thickness by FreeSurfer [36-39], and investigated potential benefits by combining these two types of feature variables. However, in this case, the number of possible feature variables we can assess tends to be much larger than the sample size - which may lead to model overfitting [40]. To tackle this, we introduced a new feature variable dimension reduction framework, in which the importance of each feature variable is quantified by a stability score for it, and then a subset of high scoring feature variables is selected for the classification task [41, 42]. Stability selection is a powerful approach to measure the quality of feature variables by evaluating how sensitive sparse linear models

(e.g., sparse logistic regression) are to each selected feature variable, where good feature variables are those more consistently selected in the final classification model, in a bootstrap process. We evaluated this framework on two independent data sets, i.e., ADNI2 [43] and NACC [44].

2. Materials and Methods:

Data:

Two independent data sets were analyzed in this study. The first dataset is from ADNI2, containing 50 HC and 112 MCI. Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). For up-to-date information, see www.adni-info.org. The second one was from NACC, containing 329 HC, 57 MCI. Demographic characteristics of the two datasets is summarized in Table 1. dMRI and T1w MRI data for each subject was analyzed. Table 2 summarizes the key data collection parameters for the two cohorts.

Feature variables:

Two types of feature variables were extracted in this study. The first type is from the gray matter using T1w MRI. FreeSurfer was used to extract 136 measurements including cortical volume and thickness for 68 brain ROIs based on Desikan-Killiany atlas [45]. The second type is from dMRI-derived structural connectome or network. The brain structural connectome was constructed using PICO [46], a whole-brain probabilistic tractography algorithm and 113 ROIs defined on the Harvard Oxford Cortical and subcortical Probabilistic Atlas [45, 47-49]. The details of computing the brain network can be referred to [35]. Each subject's network has a dimension of 113×113 , with 6,328 distinct edges connecting 113 brain ROIs (the edges are not directional and thus the network is symmetric).

Harmonization and removal of confounder effects:

There are two distinct data sets used in this study. It is likely that the imaging sequences for the two data sets may have different sensitivity to disease effects and different sources of error. Therefore, we adjusted for cohort effects, as well as age and sex. We created an indicator variable differentiating the two data sets with 1 for all subjects from ADNI2 and -1 for all subjects from NACC.

We use the commonly adopted generalized linear squares approach [50, 51] to remove the confounder effects, which assumes a variable may be linearly dependent on the confounder variables and the effects can be removed by fitting a generalized linear regression. We assume that each feature variable x has an observation value x^{obs} , given by the original feature value x^{ori} linearly biased by three confounder variables, i.e., age, sex, and cohort index, i.e.,

$$x^{obs} = w_1 \cdot age + w_2 \cdot sex + w_3 \cdot cohort + x^{ori},$$

where the coefficients can be obtained by solving a linear regression:

$$w^* = \underset{w}{\operatorname{argmin}} \sum_{i=1}^n (w^T t_i - x_i^{obs})^2,$$

where $w = [w_1, w_2, w_3]^T$ collectively denote the unknown confounder coefficients, and $t_i = (age_i, sex_i, cohort_i)$ is a vector of confounder variables for subject i , and x_i^{obs} is the observation value. After solving the linear regression, we determine the coefficients w^* and to remove the confounder effects, we thus obtain the original feature value:

$$x^{ori} = x^{obs} - (w_1 \cdot age + w_2 \cdot sex + w_3 \cdot cohort).$$

We repeat this process for all feature variables derived from both T1w MRI and dMRI. In the following analysis, we use the harmonized feature variables.

Classification modeling via sparse logistic regression:

In linear models, the sparsity means a feature variable is determined to be irrelevant if the corresponding weight is zero. Therefore, some irrelevant feature variables are discarded in the model and have no contribution to the final classification model. Sparse learning algorithms such as sparse logistic regression for classification - and LASSO [52], for linear regression - are powerful tools to build models from high dimensional data with low computational cost. The sparsity is achieved by adding sparsity-inducing regularization terms on the weight vector w such as $\lambda \|w\|_1$ to the objective function, and the final weight, or the model, is sparse with high probability. Let $x_i \in R^d$ denotes one subject where d is the number of feature variables we used, which will be elaborated later. The binary class label of this subject is denoted by $y_i \in \{-1, 1\}$, where a MCI subject is denoted as -1 and a HC subject is denoted as $+1$. Given n samples $\{\{x_1, y_1\}, \{x_2, y_2\}, \dots, \{x_n, y_n\}\}$, the loss function for the sparse logistic regression is:

$$l = \frac{1}{n} \sum_{i=1}^n \log(1 + \exp(-y_i(w^T x_i + c))) + \lambda \|w\|_1,$$

where c is the intercept, and λ is a tunable regularization parameter that is greater than or equal to 0. Here we use l_1 norm to regularize the weight vector - this will yield sparsity in the weight vector. When λ equals 0, there is no sparsity in weight vector. As λ increases, more entries in weight vector turn to 0. When λ is large enough, all the entries in weight vector become 0. By minimizing the loss function, we obtain the optimal weight vector $\hat{w}$ and intercept $\hat{c}$. For a new subject $\tilde{x}$, the probability that this subject belongs to class $\tilde{y}$ is:

$$P(\tilde{y} \mid \tilde{x}) = \frac{1}{1 + \exp(-\tilde{y}(\hat{w}^T \tilde{x} + \hat{c}))}.$$

If the probability of this subject belonging to the HC group is greater than 0.5, this subject will be labeled as HC. Otherwise this subject will be labeled as MCI.

Identify stable feature variables from brain connectome edges and T1w MRI for classification analysis:

For both the ADNI2 and NACC cohorts, the number of subjects is limited, especially when we need subjects to have both valid T1w MRI and dMRI available. When performing classification modeling, the dimension of feature variables will be much larger than the sample size for both dMRI and T1w MRI. This would lead to the “curse of dimensionality” problem where our classification models overfit training data and deliver poor generalization power. Since not all feature variables are related to the AD progression, we perform a feature variable selection procedure that ranks all the variables according to their relevance to the classification problem, and include only those feature variables in our models. There are three different feature selection models: filter models, wrapper models, and embedded models [53]. Filter models select features based on certain criteria such as Fisher score, mutual information, etc. The major disadvantage of the filter models is that they ignore the effects of the selected feature subset on the performance of the learning algorithm. Wrapper models are approaches that utilize a specific classifier to evaluate the quality of selected features. These methods test the performance of all possible subsets of the features by the classifier to determine the optimal feature subset. The computation cost of the wrapper models is quite high when the feature dimension is not small. Embedded models embed feature selection during the construction of the classifier. They have the advantage of wrapper models which include the interaction with the classification model and the advantage of filter models that are not computationally intensive. One example is sparse logistic regression. Through a theoretically sound convex programming, it eliminates features through shrinking coefficients of irrelevant features to 0. However, when noise of the level is high and sample size is extremely limited, features selected by sparse logistic regression are not stable and vulnerable to slight permutations, especially when the sample size is relatively small compared with feature dimension. Therefore, in our work, we adopt stability selection which has the advantage of the embedded models and overcomes the false selections.

Given a set of regularization parameters $\{\lambda_1, \lambda_2 \dots \lambda_n\}$, for each regularization parameter λ we obtain a set of feature variables S^λ that contribute to the final classification model in the corresponding sparse model. Stability selection is a variable selection method based on subsampling in combination with high-dimensional sparse learning algorithms. Instead of selecting one model, stability selection perturbs the data (e.g., by subsampling) many times, and we identify consistent feature variables that are included in the model, under different values of the parameter λ , across bootstrap datasets [41]. Intuitively, feature variables selected in this way are more consistently relevant to the target problem than feature variables selected only by sparse algorithms. Stability selection works as follows: we first

randomly select 50% of training samples and apply sparse logistic regression to the selected training samples with regularization parameter λ_j to build a sparse model. Let F denote the whole feature variables set and $f \in F$ denote the index of a particular feature variable in the set. The set of feature variables selected by this model is denoted by:

$$U_{\lambda_i} = \{f: w_{\lambda_i, f} \neq 0\},$$

where $w_{\lambda_i, f}$ is the weight of feature f in the model with regularization parameter λ_j . We repeat this procedure for t times. Selection probability for each feature variable is calculated as follows:

$$Pr_{f, \lambda_i} = \sum_{i=1}^t I(f \in U_{\lambda_i}) / t,$$

where $I(\cdot)$ is the indication function: $I(c) = 1$ when c is true and $I(c) = 0$ when c is false. The procedure of calculating selection probability is illustrated in the upper portion of Fig 1.

Then we vary the regularization parameter many times and calculate selection probability under these regularization parameters. By these selection probabilities, stability score for feature variable f is calculated as follows:

$$Sc(f) = \max_{\lambda_i} (Pr_{f, \lambda_i}).$$

With stability score, we can rank the variables and choose only top k stable variables, or a stability score that is larger than a pre-set threshold. The computation of the stability score is shown in the lower portion of Fig 1. After selecting feature variables by stability score, the feature dimension is reduced drastically. We will use the new feature variables set to build our model.

Classification Analysis from T1w MRI and DTI:

We use T1w MRI-derived measurements as classification variables for our base model to differentiate HC from MCI. Since the dataset is not balanced, i.e., there are more MCI than HC for ADNI2 and more HC than MCI for NACC. We apply sampling to avoid the bias introduced by the imbalance [54]. There are two sampling strategies: oversampling and subsampling. Our previous research has shown that subsampling is more effective than oversampling on AD related analysis [55]. Therefore, we randomly subsample MCI for ADNI2 and HC for NACC to make sure the training data for the two groups are balanced, i.e., the sample size of ADNI2 and NACC are the same for all three datasets we used. In this way, the classifier will be less likely biased by the group size. We randomly split the subsampled data into 90% training samples and 10% testing samples. For example, in NACC data, we have 329 HC and 57 MCI. What we have done is to sample ($\text{round}(57 * 90\%) = 51$) MCI samples and HC samples without replacement as the training data for each iteration, and use the rest as the testing data for this iteration. Similarly, for ADNI2, we sampled ($\text{round}(50 * 90\%) = 45$) HC samples and MCI samples as the training for each

iteration, and leave the rest as the testing data for that iteration. For the combined dataset, we just combined the training/testing data of ANDI2 and NACC as the training/testing data for the combined dataset. Stability selection is performed in the training phase to obtain a ranked list of T1w MRI variables and dMRI connectome edges. Sparse logistic regression is applied to build a classifier from the training samples using the top feature variables. The classifier is evaluated on the testing samples by computing AUC. We repeat experiments for 50 different random splits and compute the average AUC. To investigate if information of certain edges from dMRI can improve the performance of this base model, we propose to include the edges as added classification variables. In this paper, we include top {5, 10, ..., 45, 50} dMRI edge feature variables as additional classification variables to see how the additional dMRI information affects classification. If the classification performance is improved, then the extra information from the edges included can benefit the classification of MCI, or vice versa if the performance decreases.

3. Results

Stability selection:

We first present the results of stability selection. The regularization parameters we use are 0.01: 0.05: 0.31 for all dMRI feature variables selections and 0.01: 0.01: 0.1 for all T1w MRI feature variable selections. We repeat sparse logistic regression 1,000 times on bootstrapped datasets to calculate stability scores. Fig 2 shows the stability scores distribution for T1w MRI of three datasets. From those figures, we see stability scores decrease drastically at the beginning, and then decrease slowly till the minimal score. This means only a small set of feature variables are very stable, and there is a considerable number of feature variables that can provide a certain amount of classification information, but are not so stable. There are around 30 feature variables with stability scores larger than 0.6. Hence, the 30 feature variables can be used to train a rather stable model. Some top feature variables such as the left middle temporal thickness and volume, and the right and left entorhinal volume are also reported in many other papers [56, 57], which suggests the validity of the stability selection. In Fig 3, we present the stability selection results as brain maps.

Fig 4 shows the top 14 feature variables (only for visualization purpose) for each dataset. From the figure, we see, among the top 14 feature variables, 7 feature variables selected by ADNI2 and 8 feature variables selected by NACC are also selected by the combined dataset. The top two feature variables in the combined dataset, the cortical thickness of left rostral anterior cingulate, and its gray matter volume, are selected by both individual datasets. Also, these two feature variables have the highest stability score in the combined dataset. The common parts of NACC and ADNI2 are successfully captured by the combined dataset. This suggests that although ADNI2 and NACC are two independent cohorts, they can be successfully combined after adjusting for cohort, age, and sex effects. Since the combined dataset is larger than ADNI2 and NACC alone, models built on the combined dataset are less likely to overfit. But for the combined dataset, the gray matter volume of the left hemisphere frontal pole is not selected in either ADNI2 or NACC.

Fig 5 shows the stability score distribution for dMRI variables from the three datasets. We see the trend is similar to that of T1w MRI, with few very stable feature variables - most feature variables are consistently not important for MCI classification. We note that the key difference between T1w MRI and dMRI datasets is that there are a lot more feature variables (as there are 6,328 feature variables/connectome edges in the dMRI datasets). Robust models from sparse logistic regression will only include a very small number of variables, so there are some feature variables with 0 stability score. These feature variables do not contribute to any sparse models, even though we repeat each model for 1,000 iterations. When combine the two individual cohorts, the decreasing trend is less slow than those in individual cohorts, but it is easier to identify top ranked feature variables, as fewer stand out at the top. This also happens for the T1w MRI feature variables, meaning that variable selection can benefit from relating the two cohorts. We visualize the connectome edges with top stability scores in Fig 6.

Classification:

In this section, we present the results of classification of MCI with HC, using three datasets. For all three datasets, we first select the top 30 T1w MRI feature variables as the base model. Then we add the top {5, 10, ..., 45, 50} dMRI feature variables to the base model to see how dMRI feature variables affect classification results. We report the average AUC for each dataset. Fig 7 shows classification results for three datasets. When adding dMRI feature variables to T1w MRI, average AUC for NACC and ADNI2 increases around 5% to 15%. The p-value for model improvement is 0.0128 for ADNI2, $1.84e-17$ for NACC and $1.28e-30$ for the combined data when adding 50 dMRI feature variables. Bonferroni correction was adopted to correct for multiple comparisons (ADNI only, NACC only and Combine), thus the corrected threshold is $0.05/3 \approx 0.0167$. All our reported P values are less than this corrected threshold. This shows combination of dMRI feature variables and T1w MRI feature variables performs better than only using T1w MRI feature variables. White matter information captured by brain networks can complement the gray matter measurements from T1w MRI. Also for the combined dataset, average AUC increases consistently with NACC and ADNI2. This again shows that two cohorts can be fruitfully combined. We note that for the three datasets, when adding 10 dMRI feature variables to T1w MRI feature variables, the performance almost reaches the highest point with p-value 0.0167 for ADNI2, $4.72e-17$ for NACC and $2.92e-25$ for the combined data, which suggest those 10 dMRI feature variables are the key feature variables for classifying MCI with HC. Although there are 6,328 dMRI feature variables that provide a thorough description of brain, around 10 feature variables are sufficient to capture the major differences between MCI and HC. The top 10 feature variables for three datasets are shown in Table 3. Another observation is that the classification results of ADNI2 has higher variance than those of the rest two datasets, as illustrated in Fig. 7. One possible reason causing the variance might be the sample size: ADNI2 dataset is quite small as compared with other two datasets, i.e., NACC and the combined dataset. When the sample size is small, and the data distribution is complicated, the training data and the testing data for different iterations might differ a lot, which leads to big differences of the performance for different iterations. Therefore, for ADNI2, the variance is larger than other two datasets. We also use other two different methods to classifying MCI with HC. The two classifiers we use are SVM [58] and random forest [59].

The results are in Fig 8. We see the pattern is similar to the pattern when using logistic regression except on ADNI2 dataset. Average AUC on ADNI2 dataset first increase then decrease. That may be caused by overfitting, since the sample size of ADNI2 is too small. For both three datasets, we see adding appropriate dMRI feature variables can improve classification performance, which is consistent with the results of using logistic regression. Those results again show that white matter information together with grey matter's provide more detailed descriptive power of cognitive decline than that from grey matter alone. To make the results more reliable, we also compare the results of using other threshold of number of T1w MRI feature variables. We report the results of using top 100 T1w MRI feature variables in Fig 9. We see the results do not change much compared with the results using top 30 T1w MRI feature variables. Hence, the conclusion does not change with the threshold of number of T1w MRI feature variables.

In the last experiment, we use features selected in NACC to classify MCI and HC on ADNI2, and use features selected in ADNI to classify MCI and HC on NACC. The results are shown in Fig 10. In Fig 10, we also provide the results of using NACC selected features to classify HC and MCI on NACC and using ADNI2 selected features to classify HC and MCI on ADNI2 as a comparison. We see that for both datasets, the features selected by its own are better than the features selected by different datasets. That implies two datasets are different and the features selected by ADNI2/NACC are not the optimal features for NACC/ADNI2. Therefore, it is consistent with our observation that the high-rank features for ADNI2 and NACC are different in Fig 4.

4. Discussion

There are prior studies of AD using variable ranking and selection techniques [42, 60-64]. However, among those studies, there are significant amount of research done only on ADNI. Few studies have cross-validated the results on independent data sets, and the clinical values of the findings are limited because that the variables identified in one group may not be replicated in others. In our classification modeling process, we first ranked feature variables of T1w MRI and dMRI using their stability scores in the NACC, ADNI2, and also in the combined datasets. Then we selected the top feature variables as the key feature variables to differentiate MCI with HC on the three datasets. We see that some stable variables in independent groups no longer stand out after we combine the data. Our study brings the attention of the existence of data bias and shows evidence suggesting how it could affect the conclusion drawn from an individual cohort. Algorithms identifying key predictor variables typically depend on having a reliable dataset with a large sample size, whereas small sample sizes tend to lead to results with low confidence. By harmonizing the cohorts and combining them into a single dataset, our models are inferred from data with more subjects, so the models are more reliable. In this work, we applied generalized linear regression to harmonize two datasets. ADNI2 and NACC have many other differences including voxel size, b, the number of b0 or non-diffusion-sensitized Images. Moreover, the number of diffusion-sensitized images and other data collecting parameters such as repetition time (TR) and echo time (TE) are varying between these two cohorts. Each of these differences may contribute with a certain degree to the differences in the results derived from two cohorts [65-67]. In order to combine these two datasets to increase the sample size, we summarized

all these differences into one cohort index indicator and adopted generalized linear regression to remove all potential confound factors. However, Fig. 4 shows that among the top 14 features for ADNI2 and NACC, there are only two common features. That may suggest that even though generalized linear models are commonly used to remove cohort bias, they may be not enough to remove all the effects, especially when the effects are non-linear. Instead of seeking to remove cohort differences and use one model, one potential direction is to use domain adaptation [68-70], which assumes there are distribution differences in the cohort and learns how to align the two distributions. Domain adaptation aims to transfer models from two data sources which have different distributions. Those two domains even do not have to have homogeneous features. We plan to study these approaches in our future work.

Another challenge we addressed in this study is to do with model complexity. Model complexity is a central issue that is frequently discussed in various data analysis and machine learning models. One principle regarding model complexity is that when using a more complex model, more training samples are needed to prevent model from overfitting [71]. For commonly used linear models, the model complexity is proportional to the number of predictor variable. This has imposed a significant challenge in our analysis: when adding a large amount of dMRI derived variables to T1w MRI variables, the model complexity increases dramatically and the final classification performance is expected to drop, which may weaken the effects of added variables, or even completely override any positive effects they could add. The large amount of possible predictor variables in connectome networks and limited number of subjects even after combining cohorts can exaggerate such effects. We therefore used stability selection to rank the predictors, which has proven capability to control error rates from falsely selected variables [41, 42].

Our results show the improvement in AUC by adding dMRI is more profound in the NACC dataset - possibly because NACC participants represent more general clinical sample with vascular diseases, and where white matter integrity could play important roles in distinguishing MCI from the normals. On the other hand, ADNI is highly selective in excluding those with vascular diseases, and therefore, the additional improvement in adding dMRI is more limited in comparison with the NACC data set.

For three datasets, when adding top 10 dMRI feature variables, the performance can reach the highest AUC. MCI patients are usually suffered from the problems with memory, language, thinking and judgment in compared with normal aging [72]. The top 10 dMRI feature variables, listed in Table 3, are highly related to memory, language, thinking and judgement. In NACC, the top-ranked feature variables are concentrating on edges among the object recognition related regions (e.g. lateral Occipital Cortex [73], fusiform gyrus [74] and inferior temporal gyrus [75]) and language comprehension regions (e.g. middle and superior temporal gyrus [76]) while the top-ranked feature variables derived from ADNI2 cohort focus on the edges among the subcortical regions (e.g. insular, Amygdala, caudate and putamen). [77] reported the putamen abnormality in MCI patients while the atrophy of insular [78], amygdala [79] and Caudate [80] has been found in MCI and AD patients. Obviously, the top-ranked feature variables derived from two cohorts are not identical, which is reasonable. As we demonstrated in Table 2, there are many differences between

NACC cohort and ADNI2 cohort. Since we are using the data-driven approach, each of these differences in the cohort may have influences on the downstream feature identification, in other words, the identified disease-associated feature variables, MCI-associated feature variables in this study, from single cohort is also affected by the cohort-specific biases. One observation is in ADNI2 only, all top feature variables are in left hemisphere. To reduce these cohort-specific biases, we combined two cohorts and adopted generalized linear regression model to remove all potential confounds. By this combination, the top ranked feature variables cover the subcortical regions (e.g. insular, caudate and amygdala), which have been found when using ADNI2 only, and superior and inferior temporal gyrus, which have been found when using NACC only, as well as parahippocampal gyrus and supramarginal gyrus. Atrophy in the parahippocampal gyrus has been reported as an early biomarker of MCI and Alzheimer's disease [81] while in [82], the authors found the increased level of connectivity in correspondence of supramarginal gyrus. We believe the feature variables from the combined dataset is less affected by the noise and deserve further analyses in future studies.

The stability of selected features and models is a critical issue in data-driven approaches. In order to classify new samples and benefit clinical practice, we need to deal with the variances among different data sources and different repeats. I. Stability among different data sources: Since ADNI2 model and NACC model are quite different, and the derived features are different, directly applying any model on the new sample may not be appropriate. One possible solution is to firstly quantify the likelihood of the given new sample being in either one of the datasets i.e., comparing $P(D_{\text{ADNI2}}|x)$ and $P(D_{\text{NACC}}|x)$, where x is the new sample and D_{ADNI2} , D_{NACC} denote the events that this sample is from ADNI2, NACC, respectively. $P(D_{\text{ADNI2/NACC}}|x)$ can be calculated by Bayes' rule as follows

$$P(D_{\text{ADNI2/NACC}} | x) = \frac{P(x | D_{\text{ADNI2/NACC}})P(D_{\text{ADNI2/NACC}})}{P(x)},$$

where $P(x|D_{\text{ADNI2/NACC}})$ and $P(D_{\text{ADNI2/NACC}})$ can be estimated from the existing data in ADNI2 and NACC. These two likelihoods tell us the probability the new sample from ADNI2/NACC. Next, we can apply the classification models built on ADNI2 and NACC on this new sample to label it. Denote the label received from ADNI2 and NACC are I_{ADNI2} , I_{NACC} . The label of this sample can be estimated as $\text{sign}(I_{\text{ADNI2}} \times P(D_{\text{ADNI2}}|x) + I_{\text{NACC}} \times P(D_{\text{NACC}}|x))$. II. Stability among different repeats: (1) *Interpretability*. A set of consistent feature variables are the key to insights of underlying disease pathology. To deal with the possible inconsistent of the selected features and the models built in each iteration for one data set, one common way is to apply stability selection on the whole ADNI2/NACC data, and use the features selected by the whole ADNI2/NACC data as the feature set and build classification model on it. Then, we can use this model to classify the new patient. (2) *Predictive power*. Because of training/testing strategy used in evaluating machine learning algorithms and selecting hyper-parameters, it is quite common that the models in each repeat have different predictive power towards different samples (each model see different training samples during the training). There are two ways we can eliminate such variances. The first is to train the model using the entire data after we identify the hyper-parameters, and then it

will give us just one model for prediction with maximized predictive power, given all the data we have. The second approach to use the fact that, the models with high variances have little estimation bias (see Bias-Variance decomposition in [40]), we can perform the ensemble of these models to aggregate their advantages and produce one robust predictive model. In summary, given a new sample, we first apply the method in (II) to select features and build models for ADNI2 and NACC separately with the whole available data of the two datasets or ensemble different models. Then, we can compute the likelihood of the given new sample being in either one of the datasets and apply the labeling rule in (I) to label this new sample.

In our study, we only used a simple way to combine the two imaging modalities, i.e., concatenating the predictors from the two modalities. We expect that more sophisticated algorithms can be designed to fuse the two modalities by exploring structures within the data, and may deliver better classification performance. Our study is cross sectional. Some studies show that progression or changes in biomarkers are more closely related to the progression of clinical outcomes [83, 84] than those measured cross-sectionally. Further work is wanted to incorporate changes in biomarkers as well as to develop more sophisticated algorithms for synthesizing two modalities.

5. Conclusion

In this paper, we studied how dMRI-derived measurements help to differentiate HC subjects from those with MCI. By adding 50 dMRI feature variables to 30 T1w MRI feature variables, the AUC for classification HC with MCI is improved significantly, which indicate white matter information captured by brain networks complemented the gray matter measurements from T1w MRI in differentiating HC from individuals with MCI. Since sample size is limited, we selected feature variables by stability selection which can control the error rates of false discoveries. We performed experiments on two independent datasets, ADNI2 and NACC. Also, to cross-validate the results, we harmonized those two datasets by generalized linear squares approach, and did the same experiments on the combined datasets. The classification performance on the combined dataset is similar to that of the two independent datasets, and the top feature variables selected using three datasets are overlapped. The framework we used can also be used to classify between AD and MCI or between HC and AD. We plan to conduct this in our future work.

Acknowledgement

This study is funded in part by US Office of Naval Research (N00014-14-1-0631 and N00014-17-1-2265 to JZ), National Science Foundation (IIS-1749940, IIS-1565596 and IIS-1615597 to JZ), National Institute of Biomedical Imaging and Bioengineering (U54 EB020403 to PMT), National Institute on Aging (AG011378 and AG041851 to CRJ, P30AG053760 and P30AG008017 to HD, AG056782 to LZ). We also thank NACC staff for help with the NACC data.

The NACC database is funded by NIA/NIH Grant U01 AG016976. NACC data are contributed by the NIA funded ADCs: P30 AG019610 (PI Eric Reiman, MD), P30 AG013846 (PI Neil Kowall, MD), P50 AG008702 (PI Scott Small, MD), P50 AG025688 (PI Allan Levey, MD, PhD), P50 AG047266 (PI Todd Golde, MD, PhD), P30 AG010133 (PI Andrew Saykin, PsyD), P50 AG005146 (PI Marilyn Albert, PhD), P50 AG005134 (PI Bradley Hyman, MD, PhD), P50 AG016574 (PI Ronald Petersen, MD, PhD), P50 AG005138 (PI Mary Sano, PhD), P30 AG008051 (PI Steven Ferris, PhD), P30 AG013854 (PI M. Marsel Mesulam, MD), P30 AG008017 (PI Jeffrey Kaye, MD), P30 AG010161 (PI David Bennett, MD), P50 AG047366 (PI Victor Henderson, MD, MS), P30

AG010129 (PI Charles DeCarli, MD), P50 AG016573 (PI Frank LaFerla, PhD), P50 AG016570 (PI Marie-Francoise Chesselet, MD, PhD), P50 AG005131 (PI Douglas Galasko, MD), P50 AG023501 (PI Bruce Miller, MD), P30 AG035982 (PI Russell Swerdlow, MD), P30 AG028383 (PI Linda Van Eldik, PhD), P30 AG010124 (PI John Trojanowski, MD, PhD), P50 AG005133 (PI Oscar Lopez, MD), P50 AG005142 (PI Helena Chui, MD), P30 AG012300 (PI Roger Rosenberg, MD), P50 AG005136 (PI Thomas Montine, MD, PhD), P50 AG033514 (PI Sanjay Asthana, MD, FRCP), P50 AG005681 (PI John Morris, MD), and P50 AG047270 (PI Stephen Strittmatter, MD, PhD).

The ADNI dataset 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 (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.

Reference

- [1]. Jack CR, Jr, Knopman DS, Jagust WJ, Petersen RC, Weiner MW, Aisen PS, Shaw LM, Vemuri P, Wiste HJ, Weigand SD (2013) Update on hypothetical model of Alzheimer's disease biomarkers. *Lancet Neurol* 12, 207. [PubMed: 23332364]
- [2]. Thompson PM, Mega MS, Woods RP, Zoumalan CI, Lindshield CJ, Blanton RE, Moussai J, Holmes CJ, Cummings JL, Toga AW (2001) Cortical change in Alzheimer's disease detected with a disease-specific population-based brain atlas. *Cereb. Cortex* 11, 1–16. [PubMed: 11113031]
- [3]. Thompson PM, Hayashi KM, De Zubicaray G, Janke AL, Rose SE, Semple J, Herman D, Hong MS, Dittmer SS, Doddrell DM (2003) Dynamics of gray matter loss in Alzheimer's disease. *J Neurosci* 23, 994–1005. [PubMed: 12574429]
- [4]. Zhang Y, Dong Z, Phillips P, Wang S, Ji G, Yang J, Yuan T-F (2015) Detection of subjects and brain regions related to Alzheimer's disease using 3D MRI scans based on eigenbrain and machine learning. *Front Comput Neurosci* 9, 66. [PubMed: 26082713]
- [5]. Zhang Y, Wang S (2015) Detection of Alzheimer's disease by displacement field and machine learning. *PeerJ* 3, e1251. [PubMed: 26401461]
- [6]. Wang S, Zhang Y, Liu G, Phillips P, Yuan T-F (2016) Detection of Alzheimer's disease by three-dimensional displacement field estimation in structural magnetic resonance imaging. *J Alzheimers Dis* 50, 233–248. [PubMed: 26682696]
- [7]. Escudero J, Zajicek JP, Ifeachor E (2011) Machine Learning classification of MRI features of Alzheimer's disease and mild cognitive impairment subjects to reduce the sample size in clinical trials. *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE IEEE*, pp. 7957–7960.
- [8]. Ota K, Oishi N, Ito K, Fukuyama H, SEAD-J Study Group (2014) A comparison of three brain atlases for MCI prediction. *J Neurosci Methods* 221, 139–150. [PubMed: 24140118]
- [9]. Zhou J, Liu J, Narayan VA, Ye J, Initiative AsDN (2013) Modeling disease progression via multi-task learning. *Neuroimage* 78, 233–248. [PubMed: 23583359]
- [10]. Kolibas E, Korinkova V, Novotny V, Vajdickova K, Hunakova D (2000) ADAS-Cog (Alzheimer's Disease Assessment Scale-cognitive subscale)-validation of the Slovak version. *Bratisl Lek Listy* 101, 598–602. [PubMed: 11218956]
- [11]. Lin F, Weng S, Xie B, Wu G, Lei H (2011) Abnormal frontal cortex white matter connections in bipolar disorder: a DTI tractography study. *J Affect Disord* 131, 299–306. [PubMed: 21236494]

- [12]. Arienzo D, Leow A, Brown JA, Zhan L, GadElkarim J, Hovav S, Feusner JD (2013) Abnormal brain network organization in body dysmorphic disorder. *Neuropsychopharmacology* 38, 1130–1139. [PubMed: 23322186]
- [13]. Haney-Caron E, Caprihan A, Stevens MC (2014) DTI-measured white matter abnormalities in adolescents with Conduct Disorder. *J Psychiatr Res* 48, 111–120. [PubMed: 24139595]
- [14]. Zhou Y, Dougherty JH, Hubner KF, Bai B, Cannon RL, Hutson RK (2008) Abnormal connectivity in the posterior cingulate and hippocampus in early Alzheimer's disease and mild cognitive impairment. *Alzheimers Dement* 4, 265–270. [PubMed: 18631977]
- [15]. Wang Q, Zhan L, Thompson PM, Dodge HH, Zhou J (2016) Discriminative fusion of multiple brain networks for early mild cognitive impairment detection. In *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on IEEE*, pp. 568–572.
- [16]. Daianu M, Dennis EL, Jahanshad N, Nir TM, Toga AW, Jack CR, Weiner MW, Thompson PM (2013) Alzheimer's disease disrupts rich club organization in brain connectivity networks. In *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on IEEE*, pp. 266–269.
- [17]. Ajilore O, Vizueta N, Walshaw P, Zhan L, Leow A, Altshuler LL (2015) Connectome signatures of neurocognitive abnormalities in euthymic bipolar I disorder. *J Psychiatr Res* 68, 37–44. [PubMed: 26228398]
- [18]. Zhan L, Liu Y, Wang Y, Zhou J, Jahanshad N, Ye J, Thompson PM (2015) Boosting brain connectome classification accuracy in Alzheimer's disease using higher-order singular value decomposition. *Front Neurosci* 9, 257. [PubMed: 26257601]
- [19]. Wang T, Shi F, Jin Y, Yap P-T, Wee C-Y, Zhang J, Yang C, Li X, Xiao S, Shen D (2016) Multilevel deficiency of white matter connectivity networks in Alzheimer's disease: a diffusion MRI study with DTI and HARDI models. *Neural Plast* 2016.
- [20]. Zhan L, Nie Z, Ye J, Wang Y, Jin Y, Jahanshad N, Prasad G, de Zubicaray G, McMahon K, Martin N (2014) Multiple stages classification of Alzheimer's disease based on structural brain networks using generalized low rank approximations (GLRAM) In *Comput Diffus MRI Springer*, pp. 35–44.
- [21]. Zarei M, Patenaude B, Damoiseaux J, Morgese C, Smith S, Matthews PM, Barkhof F, Rombouts S, Sanz-Arigita E, Jenkinson M (2010) Combining shape and connectivity analysis: an MRI study of thalamic degeneration in Alzheimer's disease. *Neuroimage* 49, 1–8. [PubMed: 19744568]
- [22]. Zhang D, Wang Y, Zhou L, Yuan H, Shen D, Initiative AsDN (2011) Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage* 55, 856–867. [PubMed: 21236349]
- [23]. Möller C, Hafkemeijer A, Pijnenburg YA, Rombouts SA, van der Grond J, Dopper E, van Swieten J, Versteeg A, Pouwels PJ, Barkhof F (2015) Joint assessment of white matter integrity, cortical and subcortical atrophy to distinguish AD from behavioral variant FTD: A two-center study. *Neuroimage Clin* 9, 418–429. [PubMed: 26594624]
- [24]. Schouten TM, Koini M, de Vos F, Seiler S, van der Grond J, Lechner A, Hafkemeijer A, Möller C, Schmidt R, de Rooij M (2016) Combining anatomical, diffusion, and resting state functional magnetic resonance imaging for individual classification of mild and moderate Alzheimer's disease. *Neuroimage Clin* 11, 46–51. [PubMed: 26909327]
- [25]. Zu C, Jie B, Liu M, Chen S, Shen D, Zhang D, Initiative AsDN (2016) Label-aligned multi-task feature learning for multimodal classification of Alzheimer's disease and mild cognitive impairment. *Brain Imaging Behav* 10, 1148–1159. [PubMed: 26572145]
- [26]. Wang J, Wang Q, Peng J, Nie D, Zhao F, Kim M, Zhang H, Wee CY, Wang S, Shen D (2017) Multi-task diagnosis for autism spectrum disorders using multi-modality features: A multi-center study. *Hum Brain Mapp* 38, 3081–3097. [PubMed: 28345269]
- [27]. Verstraete E, Polders DL, Mandl RC, Van Den Heuvel MP, Veldink JH, Luijten P, Van Den Berg LH, Hoogduin J (2014) Multimodal tract-based analysis in ALS patients at 7T: a specific white matter profile? *Amyotroph Lateral Scler Frontotemporal Degener* 15, 84–92. [PubMed: 24325276]
- [28]. Wang L, Goldstein FC, Veledar E, Levey AI, Lah JJ, Meltzer CC, Holder CA, Mao H (2009) Alterations in cortical thickness and white matter integrity in mild cognitive impairment

measured by whole-brain cortical thickness mapping and diffusion tensor imaging. *AJNR Am J Neuroradiol* 30, 893–899. [PubMed: 19279272]

- [29]. Zhang Y, Schuff N, Camacho M, Chao LL, Fletcher TP, Yaffe K, Woolley SC, Madison C, Rosen HJ, Miller BL (2013) MRI markers for mild cognitive impairment: comparisons between white matter integrity and gray matter volume measurements. *PloS one* 8, e66367. [PubMed: 23762488]
- [30]. Cherubini A, Péran P, Spoletini I, Di Paola M, Di Iulio F, Hagberg GE, Sancesario G, Gianni W, Bossu P, Caltagirone C (2010) Combined volumetry and DTI in subcortical structures of mild cognitive impairment and Alzheimer's disease patients. *J Alzheimers Dis* 19, 1273–1282. [PubMed: 20308792]
- [31]. Müller MJ, Greverus D, Dellani PR, Weibrich C, Wille PR, Scheurich A, Stoeter P, Fellgiebel A (2005) Functional implications of hippocampal volume and diffusivity in mild cognitive impairment. *Neuroimage* 28, 1033–1042. [PubMed: 16084115]
- [32]. Müller MJ, Greverus D, Weibrich C, Dellani PR, Scheurich A, Stoeter P, Fellgiebel A (2007) Diagnostic utility of hippocampal size and mean diffusivity in amnesic MCI. *Neurobiol Aging* 28, 398–403. [PubMed: 16529847]
- [33]. Gao Y, Wee C-Y, Kim M, Giannakopoulos P, Montandon M-L, Haller S, Shen D (2015) MCI identification by joint learning on multiple MRI data In *International Conference on Medical Image Computing and Computer-Assisted Intervention* Springer, pp. 78–85.
- [34]. Palesi F, Vitali P, Chiarati P, Castellazzi G, Caverzasi E, Pichiecchio A, Colli-Tibaldi E, D'Amore F, D'Errico I, Sinforiani E (2012) DTI and MR volumetry of hippocampus-PC/PCC circuit: in search of early micro-and macrostructural signs of Alzheimer's disease *Neurol Res Int* 2012.
- [35]. Zhan L, Zhou J, Wang Y, Jin Y, Jahanshad N, Prasad G, Nir TM, Leonardo CD, Ye J, Thompson PM (2015) Comparison of nine tractography algorithms for detecting abnormal structural brain networks in Alzheimer's disease. *Front Aging Neurosci.* 7.
- [36]. Desikan RS, Cabral HJ, Hess CP, Dillon WP, Glastonbury CM, Weiner MW, Schmansky NJ, Greve DN, Salat DH, Buckner RL (2009) Automated MRI measures identify individuals with mild cognitive impairment and Alzheimer's disease. *Brain* 132, 2048–2057. [PubMed: 19460794]
- [37]. Bernal-Rusiel JL, Reuter M, Greve DN, Fischl B, Sabuncu MR, Initiative AsDN (2013) Spatiotemporal linear mixed effects modeling for the mass-univariate analysis of longitudinal neuroimage data. *Neuroimage* 81, 358–370. [PubMed: 23702413]
- [38]. Schmitter D, Roche A, Maréchal B, Ribes D, Abdulkadir A, Bach-Cuadra M, Daducci A, Granziera C, Klöppel S, Maeder P (2015) An evaluation of volume-based morphometry for prediction of mild cognitive impairment and Alzheimer's disease. *Neuroimage Clin* 7, 7–17. [PubMed: 25429357]
- [39]. Hirjak D, Huber M, Kirchler E, Kubera KM, Karner M, Sambataro F, Freudenmann RW, Wolf RC (2017) Cortical features of distinct developmental trajectories in patients with delusional infestation. *Prog Neuropsychopharmacol Biol Psychiatry* 76, 72–79. [PubMed: 28257853]
- [40]. Bishop C (2007) *Pattern Recognition and Machine Learning (Information Science and Statistics)*, 1st edn. 2006. corr. 2nd printing edn.
- [41]. Meinshausen N, Bühlmann P (2010) Stability selection. *J R Stat Soc Series B Stat Methodol* 72, 417–473.
- [42]. Ye J, Farnum M, Yang E, Verbeeck R, Lobanov V, Raghavan N, Novak G, DiBernardo A, Narayan VA (2012) Sparse learning and stability selection for predicting MCI to AD conversion using baseline ADNI data. *BMC Neurol* 12, 46. [PubMed: 22731740]
- [43]. Jack CR, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, Borowski B, Britson PJ, L Whitwell J, Ward C (2008) The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *J Magn Reson Imaging* 27, 685–691. [PubMed: 18302232]
- [44]. Beekly DL, Ramos EM, Lee WW, Deitrich WD, Jacka ME, Wu J, Hubbard JL, Koepsell TD, Morris JC, Kukull WA (2007) The National Alzheimer's Coordinating Center (NACC) database: the uniform data set. *Alzheimer Dis Assoc Disord* 21, 249–258. [PubMed: 17804958]
- [45]. Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT (2006) An automated labeling system for subdividing the human

cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31, 968–980. [PubMed: 16530430]

- [46]. Parker GJ, Haroon HA, Wheeler-Kingshott CA (2003) A framework for a streamline-based probabilistic index of connectivity (PICO) using a structural interpretation of MRI diffusion measurements. *J Magn Reson Imaging* 18, 242–254. [PubMed: 12884338]
- [47]. Frazier JA, Chiu S, Breeze JL, Makris N, Lange N, Kennedy DN, Herbert MR, Bent EK, Koneru VK, Dieterich ME (2005) Structural brain magnetic resonance imaging of limbic and thalamic volumes in pediatric bipolar disorder. *Am J Psychiatry* 162, 1256–1265. [PubMed: 15994707]
- [48]. Makris N, Goldstein JM, Kennedy D, Hodge SM, Caviness VS, Faraone SV, Tsuang MT, Seidman LJ (2006) Decreased volume of left and total anterior insular lobule in schizophrenia. *Schizophr Res* 83, 155–171. [PubMed: 16448806]
- [49]. Goldstein JM, Seidman LJ, Makris N, Ahern T, O'Brien LM, Caviness VS, Kennedy DN, Faraone SV, Tsuang MT (2007) Hypothalamic abnormalities in schizophrenia: sex effects and genetic vulnerability. *Biol. Psychiatry* 61, 935–945. [PubMed: 17046727]
- [50]. McCullagh P (1984) Generalized linear models. *Eur J Oper Res* 16, 285–292.
- [51]. McNamee R (2005) Regression modelling and other methods to control confounding. *Occup. Environ. Med* 62, 500–506. [PubMed: 15961628]
- [52]. Tibshirani R (1996) Regression shrinkage and selection via the lasso. *J R Stat Soc Series B Stat Methodol*, 267–288.
- [53]. Tang J, Alelyani S, Liu H (2014) Feature selection for classification: A review. *Data Classification: Algorithms and Applications*, 37.
- [54]. Chawla NV (2009) Data mining for imbalanced datasets: An overview In *Data mining and knowledge discovery handbook* Springer, pp. 875–886.
- [55]. Dubey R, Zhou J, Wang Y, Thompson PM, Ye J, Initiative AsDN (2014) Analysis of sampling techniques for imbalanced data: An n= 648 ADNI study. *NeuroImage* 87, 220–241. [PubMed: 24176869]
- [56]. Juottonen K, Laakso M, Insausti R, Lehtovirta M, Pitkänen A, Partanen K, Soininen H (1998) Volumes of the entorhinal and perirhinal cortices in Alzheimer's disease. *Neurobiol Aging* 19, 15–22. [PubMed: 9562498]
- [57]. Dickerson BC, Feczko E, Augustinack JC, Pacheco J, Morris JC, Fischl B, Buckner RL (2009) Differential effects of aging and Alzheimer's disease on medial temporal lobe cortical thickness and surface area. *Neurobiol Aging* 30, 432–440. [PubMed: 17869384]
- [58]. Chang C-C, Lin C-J (2011) LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol (TIST)* 2, 27.
- [59]. Liaw A, Wiener M (2002) Classification and regression by randomForest. *R news* 2, 18–22.
- [60]. Chaves R, Ramírez J, Górriz J, López M, Salas-Gonzalez D, Alvarez I, Segovia F (2009) SVM-based computer-aided diagnosis of the Alzheimer's disease using t-test NMSE feature selection with feature correlation weighting. *Neurosci Lett* 461, 293–297. [PubMed: 19549559]
- [61]. Batmanghelich KN, Dong HY, Pohl KM, Taskar B, Davatzikos C (2011) Disease classification and prediction via semi-supervised dimensionality reduction. In *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on IEEE*, pp. 1086–1090.
- [62]. Chu C, Hsu A-L, Chou K-H, Bandettini P, Lin C, Initiative AsDN (2012) Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. *Neuroimage* 60, 59–70. [PubMed: 22166797]
- [63]. Varol E, Gaonkar B, Erus G, Schultz R, Davatzikos C (2012) Feature ranking based nested support vector machine ensemble for medical image classification. In *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on IEEE*, pp. 146–149.
- [64]. Beheshti I, Demirel H, Initiative AsDN (2016) Feature-ranking-based Alzheimer's disease classification from structural MRI. *Magn Reson Imaging* 34, 252–263. [PubMed: 26657976]
- [65]. Jahanshad N, Zhan L, Bernstein MA, Borowski BJ, Jack CR, Toga AW, Thompson PM (2010) Diffusion tensor imaging in seven minutes: determining trade-offs between spatial and directional resolution. In *Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on IEEE*, pp. 1161–1164.

- [66]. Zhan L, Franc D, Patel V, Jahanshad N, Jin Y, Mueller BA, Bernstein MA, Borowski BJ, Jack CR, Toga AW (2012) How do spatial and angular resolution affect brain connectivity maps from diffusion MRI?. In Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on IEEE, pp. 1–4.
- [67]. Zhan L, Jahanshad N, Ennis DB, Jin Y, Bernstein MA, Borowski BJ, Jack CR, Toga AW, Leow AD, Thompson PM (2013) Angular versus spatial resolution trade-offs for diffusion imaging under time constraints. *Hum Brain Mapp* 34, 2688–2706. [PubMed: 22522814]
- [68]. Kulis B, Saenko K, Darrell T (2011) What you saw is not what you get: Domain adaptation using asymmetric kernel transforms. In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on IEEE, pp. 1785–1792.
- [69]. Duan L, Tsang IW, Xu D, Chua T-S (2009) Domain adaptation from multiple sources via auxiliary classifiers. In Proceedings of the 26th Annual International Conference on Machine Learning ACM, pp. 289–296.
- [70]. Xiao M, Guo Y (2015) Feature space independent semi-supervised domain adaptation via kernel matching. *IEEE Trans Pattern Anal Mach Intell* 37, 54–66. [PubMed: 26353208]
- [71]. Myung IJ (2000) The importance of complexity in model selection. *J Math Psychol* 44, 190–204. [PubMed: 10733864]
- [72]. Petersen RC, Negash S (2008) Mild cognitive impairment: an overview. *CNS Spectr* 13, 45–53.
- [73]. Grill-Spector K, Kourtzi Z, Kanwisher N (2001) The lateral occipital complex and its role in object recognition. *Vision Res* 41, 1409–1422. [PubMed: 11322983]
- [74]. Grill-Spector K, Sayres R (2008) Object recognition: Insights from advances in fMRI methods. *Curr Dir Psychol Sci* 17, 73–79.
- [75]. Gerlach C, Aaside C, Humphreys G, Gade A, Paulson OB, Law I (2002) Brain activity related to integrative processes in visual object recognition: bottom-up integration and the modulatory influence of stored knowledge. *Neuropsychologia* 40, 1254–1267. [PubMed: 11931928]
- [76]. Onitsuka T, Shenton ME, Salisbury DF, Dickey CC, Kasai K, Toner SK, Frumin M, Kikinis R, Jolesz FA, McCarley RW (2004) Middle and inferior temporal gyrus gray matter volume abnormalities in chronic schizophrenia: an MRI study. *Am J Psychiatry* 161, 1603–1611. [PubMed: 15337650]
- [77]. Cao Q, Jiang K, Zhang M, Liu Y, Xiao S, Zuo C, Huang H (2003) Brain glucose metabolism and neuropsychological test in patients with mild cognitive impairment. *Chin Med J* 116, 1235–1238. [PubMed: 12935418]
- [78]. Sluimer JD, van der Flier WM, Karas GB, van Schijndel R, Barnes J, Boyes RG, Cover KS, Olabarriaga SD, Fox NC, Scheltens P (2009) Accelerating regional atrophy rates in the progression from normal aging to Alzheimer's disease. *Eur Radiol* 19, 2826. [PubMed: 19618189]
- [79]. Miller MI, Younes L, Ratnanather JT, Brown T, Reigel T, Trinh H, Tang X, Barker P, Mori S, Albert M (2012) Amygdala atrophy in MCI/Alzheimer's disease in the BIOCARD cohort based on diffeomorphic morphometry. In Medical image computing and computer-assisted intervention: MICCAI... International Conference on Medical Image Computing and Computer-Assisted Intervention NIH Public Access, p. 155.
- [80]. Madsen SK, Ho AJ, Hua X, Saharan PS, Toga AW, Jack C, Weiner MW, Thompson PM, Initiative AsDN (2010) 3D maps localize caudate nucleus atrophy in 400 Alzheimer's disease, mild cognitive impairment, and healthy elderly subjects. *Neurobiol Aging* 31, 1312–1325. [PubMed: 20538376]
- [81]. Echavarri C, Aalten P, Uylings HB, Jacobs H, Visser P, Gronenschild E, Verhey F, Burgmans S (2011) Atrophy in the parahippocampal gyrus as an early biomarker of Alzheimer's disease. *Brain Struct Funct* 215, 265–271. [PubMed: 20957494]
- [82]. Esposito R, Mosca A, Pieramico V, Cieri F, Cera N, Sensi SL (2013) Characterization of resting state activity in MCI individuals. *PeerJ* 1, e135. [PubMed: 24010015]
- [83]. Dodge HH, Zhu J, Harvey D, Saito N, Silbert LC, Kaye JA, Koeppe RA, Albin RL, Initiative AsDN (2014) Biomarker progressions explain higher variability in stage-specific cognitive decline than baseline values in Alzheimer disease. *Alzheimers Dement* 10, 690–703. [PubMed: 25022534]

- [84]. Silbert LC, Howieson DB, Dodge H, Kaye JA (2009) Cognitive impairment risk White matter hyperintensity progression matters. *Neurology* 73, 120–125. [PubMed: 19597134]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig 1.

The pipeline of computing the stability score. The warmer color indicates a higher probability of selection. (a) Calculating the selection probability using different regularization parameters. (b) illustration of using selection probability to calculate stability score.

Fig 2.

The performance of T1w stability scores vs. the number of T1w feature variables from three datasets (ADNI2 only, NACC only and the combined dataset).

Fig 3. Brain surface maps showing the spatial distribution of the T1w MRI stability scores for ADNI2 only, NACC only and the combined dataset.

Fig 4.
Top 14 feature variables selected by three datasets for T1w MRI.

Fig5. Distribution of dMRI stability scores for three datasets (ADNI2 only, NACC only and the combined dataset).

Fig 6.
Illustration of Connectome edges with top stability scores (only top 10 showed here).

Fig 7. Classification performance for MCI vs. HC using sparse logistic regression. The x-axis represents the number of dMRI feature variables included in the T1w MRI feature variables.

Fig 8. Classification performance for HC vs. MCI using SVM and random forest. The x-axis represents the number of dMRI feature variables included in the T1w MRI feature variables.

Fig 9. Classification performance for MCI vs. HC when using 100 T1w MRI feature variables to build base model.

Fig. 10.

Classification results of using features selected in NACC to classify MCI and HC on ADNI2, and using features selected in ADNI2 to classify MCI and HC on NACC. The lines with triangle markers are the results of using NACC selected features to classify HC and MCI on NACC and using ADNI2 selected features to classify HC and MCI on ADNI2 as a comparison.

Table 1.

Demographic information for the two cohorts (ADNI2 and NACC). The p -values for the difference between ADNI2 and NACC are 0.023 for sex and 3.88e-23 for age. The last column is the p -value for the difference between MCI and HC.

ADNI2	HC	MCI	Total	p -value
Number	50	112	162	-
age	69.36 $\pm$ 15.40	71.68 $\pm$ 9.93	70.96 $\pm$ 11.89	0.0016
sex	22M / 28F	71M / 41F	93M / 69F	0.0040
NACC	HC	MCI	Total	-
Number	329	57	386	-
age	60.96 $\pm$ 8.96	73.60 $\pm$ 7.93	63.82 $\pm$ 9.73	0.0100
sex	107M / 222F	38M / 19F	145M / 241F	0.0046

Table 2.

Parameters for dMRI and T1w MRI data for ADNI2 and NACC.

	ADNI2	NACC
b value	1000 s/mm ²	1300 s/mm ²
Number of b0 images	5	8
Number of diffusion weighted images	41	40
T1w MRI voxel size	1.0156*1.0156*1.2 mm ³	1.0*1.0*1.2 mm ³
T1w MRI TR	6.98ms	8.16ms
T1w MRI TE	2.85ms	3.18ms
T1w MRI Image dimension	256*256*196	256*256*156
dMRI voxel size	2.7*2.7*2.7 mm ³	0.94*0.94*2.9 mm ³
dMRI TR	9050 ms	8000 ms
dMRI TE	Minimum	81.8 ms
dMRI Image dimension	128*128*59	256*256*52

Table 3.

Top 10 dMRI feature variables identified using three datasets. Each feature variable represents the connection between ROI1 and ROI2.

ADNI2	ROI1	ROI2
1	Left Caudate	Left Pallidum
2	Left Insular Cortex	Left Central Opercular Cortex
3	Left Superior Parietal Lobule	Left Lateral Occipital Cortex, Superior Division
4	Left Amygdala	Left Central Opercular Cortex
5	Left Pallidum	Left Insular Cortex
6	Brainstem	Left Insular Cortex
7	Left Pallidum	Left Central Opercular Cortex
8	Left Insular Cortex	Left Precentral Gyrus
9	Left Caudate	Left Putamen
10	Left Precentral Gyrus	Left Juxtapositional Lobule Cortex
NACC	ROI1	ROI2
1	Right Pallidum	Left Lateral Occipital Cortex, Inferior Division
2	Right Lateral Occipital Cortex, Inferior Division	Left Temporal Occipital Fusiform Cortex
3	Left Middle Temporal Gyrus, Posterior Division	Right Occipital Pole
4	Right Superior Temporal Gyrus, Posterior Division	Left Supramarginal Gyrus, Posterior Division
5	Left Middle Temporal Gyrus, Posterior Division	Right Supracalcarine Cortex
6	Right Inferior Temporal Gyrus, Anterior Division	Right Occipital Fusiform Cortex'
7	Left Heschl'S Gyrus	Left Cerebellum
8	Right Parietal Opercular Cortex	Right Cerebellum
9	Left Parietal Opercular Cortex	LeftOccipitalPole
10	Right Accumbens	Left Inferior Temporal Gyrus, Temporooccipital Part'
Combined	ROI1	ROI2
1	Right Parahippocampal Gyrus, Posterior Division	Right Heschl'S Gyrus
2	Right Amygdala	Left Cerebellum
3	Right Inferior Temporal Gyrus, Temporooccipital Part	Right Supramarginal Gyrus, Anterior Division
4	Brainstem	Left Insular Cortex
5	Left Insular Cortex	Left Frontal Opercular Cortex
6	Right Superior Temporal Gyrus, Posterior Division	Left Supramarginal Gyrus, Posterior Division
7	Left Caudate	Left Pallidum
8	Left Frontal Pole	Left Frontal Opercular Cortex
9	Left Inferior Temporal Gyrus, Temporooccipital Part	Right Supramarginal Gyrus, Posterior Division
10	Right Superior Parietal Lobule	Left Planum Temporale