Extracting Patterns of Morphometry Distinguishing HIV Associated Neurodegeneration from Mild Cognitive Impairment via Group Cardinality Constrained Classification

Yong Zhang[1,*], **Dongjin Kwon**[1,2], **Pardis Esmaeili-Firidouni**[3], **Adolf Pfefferbaum**[2], **Edith V. Sullivan**[1], **Harold Javitz**[4], **Victor Valcour**[3], and **Kilian M. Pohl**[1,2]

[1]Department of Psychiatry & Behavioral Sciences, Stanford University, Stanford, CA 94305 USA

[2]Center for Health Sciences, SRI International, Menlo Park, CA, 94025 USA

[3]Memory and Aging Center, UCSF, San Francisco, CA, USA

[4]Center for Technology in Learning in the Education Division, SRI International, Menlo Park, CA, 94025 USA

Abstract

HIV-Associated Neurocognitive Disorder (HAND) is the most common constellation of cognitive dysfunctions in chronic HIV infected patients age 60 or older in the U.S. Only few published methods assist in distinguishing HAND from other forms of age-associated cognitive decline, such as Mild Cognitive Impairment (MCI). In this report, we proposed a data-driven, nonparameteric model to identify morphometric patterns separating HAND from MCI due to non-HIV conditions in this older age group. This model enhanced the potential for group separation by combining a smaller, longitudinal data set containing HAND samples with a larger, public data set including MCI cases. Using cross-validation, we trained a linear model on healthy controls to harmonize the volumetric scores extracted from MRIs for demographic and acquisition differences between the two independent, disease-specific data sets. Next, we identified patterns distinguishing HAND from MCI via a group sparsity constrained logistic classifier. Unlike existing approaches, our classifier directly solved the underlying minimization problem by decoupling the minimization of the logistic regression function from enforcing the group sparsity constraint. The extracted patterns consisted of eight regions that distinguished HAND from MCI on a significant level while being indifferent to differences in demographics and acquisition between the two sets. Individually selecting regions through conventional morphometric group analysis resulted in a larger number of regions that were less accurate. In conclusion, simultaneously analyzing all brain regions and time points for disease specific patterns contributed to distinguishing with high accuracy HAND-related impairment from cognitive impairment found in the HIV uninfected, MCI cohort.

---

*Corresponding Author: Yong Zhang, yzhang83@stanford.edu, 1-408-859-2389, Department of Psychiatry & Behavioral Sciences, Stanford University, 450 Serra Mall, Stanford, CA 94305.

Keywords

Sparsity; Logistic Regression; HIV-Associated Neurocognitive Disorder; Mild Cognitive Impairment; MRI

Introduction

HIV-Associated Neurocognitive Disorder (HAND) affects upwards of 50% of HIV infected (HIV+) individuals [Heaton et al. (2011); McArthur et al. (2010)] and is linked to decreased everyday function, reduced adherence to therapy, and increased mortality [Heaton et al. (2011); Ettenhofer et al. (2009); Heaton et al. (2004)]. Distinguishing HAND from other forms of age-associated cognitive compromise, such as those broadly termed Mild Cognitive Impairment (MCI) and often due to Alzheimer's disease [Brew et al. (2009); Cohen et al. (2015)], is an emerging clinical dilemma as the neuropathogenic mechanisms associated with HIV and aging are mostly unknown. The success of combination antiretroviral therapy (cART) has led to a rapid rise in the number of chronic HIV+ patients over the age of 60 in the U.S. [Kirk and Goetz (2009)]. Aging HAND patients are at risk of also developing age-related cognitive decline due to Alzheimer's disease. However, clinicians currently do not have assessment tools to perform a differential diagnosis on HAND patients for early treatment of age-associated cognitive impairments.

To enhance separation of these overlapping conditions, an increasing number of neuroimaging studies have been investigating the independent and interactive effects of HIV and aging [Towgood et al. (2012); Chang et al. (2013); Cysique et al. (2013); Thomas et al. (2013); Liu and Press (2013); Ances and Hammoud (2014); Pfefferbaum et al. (2014); Holt et al. (2012); Cysique and Brew (2014)]. However, imaging studies published on aging with chronic HIV infection limited analysis to parametric models assuming that groups comprising subjects with similar demographics are homogeneous [Cysique et al. (2013); Thomas et al. (2013); Liu and Press (2013); Ances et al. (2012); Chang et al. (2011); Cardenas et al. (2009); Jahanshad et al. (2012); McMurtray et al. (2008); Nir et al. (2014); Sullivan and Pfefferbaum (2003)]. The homogeneity of variance assumption was generally not met due to the notable variation of brain anatomy even in the normal population[Pfefferbaum et al. (2013)]; thus most ongoing investigations are limited to small sizes of carefully selected samples. In addition, earlier studies were restricted to individuals under the age of 60 limiting the relevance of the findings to older cohorts, who are at increased risk for other neurodegenerative disorders [Valcour et al. (2004)]. To identify morphometric patterns distinguishing HAND from MCI in this older age group, we now propose a data-driven, nonparameteric model that enhances the potential for discovery by complementing a smaller, longitudinal MRI data set containing HAND samples from the UCSF HIV Over 60 Cohort Study (UHES) with a larger, publicly available data set from the Alzheimer's Disease Neuroimaging Initiative (ADNI), which includes MCI samples.

Identifying morphometric group differences based on independently collected MR data sets were subject to differences in MR acquisition, clinical scoring, and demography characteristics [Han et al. (2006); Stonnington et al. (2008); Focke et al. (2011); Simmons et

al. (2011); Jovicich et al. (2013); Westman et al. (2011)]. Commonly, a general additive model has been applied to the morphometric measurements to regress out the cohort specific factors [Bartsch et al. (2014); Deoni et al. (2008); van Erp et al. (2015); Pfefferbaum et al. (2015)]. For each region, the residual scores were then applied to statistical tests to investigate morphometric group differences, such as in [Pfefferbaum et al. (2013)]. To refine analysis on an individual subject level, the medical imaging community has frequently relied on sparse classifiers [Xiang et al. (2014); Rosa et al. (2015)]. From the large set of image features, sparse classifiers pick a few scores informative with respect to distinguishing cohorts under investigation. This approach enabled extraction of disease specific patterns from a relatively small number of samples [Ye et al. (2012); Liu et al. (2012)], *i.e.*, the number of samples was much smaller than the number of measurements extracted from the images.

The selection process of sparse classifiers counts the number of features deemed as influential via the l_0-"norm" and then determines a solution so that the count is below a predefined threshold [Yamashita et al. (2008); Carroll et al. (2009); Rao et al. (2011); Lv et al. (2015)]. A generalization of this concept are group sparsity models, which enforce sparsity on groups of image features [Ng et al. (2010); Wu et al. (2010); Ryali et al. (2010)]. Grouping can be used to model application specific constraints, such as the similar effect of MCI and HAND on each cerebral hemisphere. To solve the underlying minimization problem, however, these methods relaxed the feature selection process from (group) cardinality constraints to weighting features by replacing, for example, the l_0-"norm" with the l_1-norm or in the case of group cardinality constraints with the l_2-norm [Meier et al. (2008); Friedman et al. (2010); Li et al. (2012); Yuan et al. (2012)]. The solution of those methods related to the original sparse problem only under specific assumptions, e.g., compressed sensing [Candès et al. (2006)], which generally do not hold for medical image applications. In addition, the number of measures selected by the classifier depended on the training data due to the soft selection scheme. Confining selection to a predefined number required choosing measures whose weight was above a certain threshold. This practice, however, has been controversial because classification depends on the measures below the threshold, and the relevance of those weights with respect to the effects of disease on morphometric features were unclear [Haufe et al. (2014); Sabuncu (2014)]. Alternatively, the upper bound associated with the sparse constraint was tuned with respect to a training data set so the classifier returned the wanted number of measures [Vounou et al. (2012); Zhang and Shen (2012); Ma and Huang (2008); Zhang et al. (2012)]. The tuning, however, was data dependent, *i.e.*, the upper bound generally needed to be adjusted for each training set to enforce the same number of scores to be selected. Comparing patterns of different subsets of even the same data set, *i.e.*, folds, is thus not trivial as each pattern is the solution to a minimization problem, whose sparsity constraint is unique to a fold. Our approach [Zhang et al. (2016)] avoided these issues by working on the original, group sparsity constrained, logistic classification model defined by the the l_0-"norm". By determining a local solution to this non-convex and non-continuous minimization problem, our method directly selected patterns (without thresholding or changing upper bounds) that were potentially meaningful image markers for distinguishing diseases. In [Zhang et al. (2016)], we compared the accuracy of our algorithm to a model constrained by (non-group) sparsity, and an approach

that relaxed sparsity constraints via the l_1- or l_2-norm and then determined the optimal global solutions for the relaxed model. All methods were applied to a data set containing a ground-truth diagnosis, *i.e.*, the cine MRI images of adults with and without reconstructive surgery of tetralogy of Fallot (TOF) during infancy. In accordance with the cardiac literature, our method correctly identified the right ventricle to be most impacted by TOF and generally obtained statistically significant higher classification accuracy than the alternative approaches.

To indentify image markers destinguishing MCI from HAND in patients of age 60 and older, this article describes our findings when applying our group sparsity constraint, logistic classifier to a dataset, which combined the UHES data set containing aging HAND cases (n = 15) and a matched healthy control group (n = 21) with a larger, publicly available, and demographically matched subset of the ADNI project consisting of MCI cases (n = 80) and healthy controls (n = 26) (see also Figure 1). Each sample was characterized by the region-of-interest (ROI) specific volume scores of baseline and follow-up. As the data associated with each disease were collected in separate studies, we regressed out demographic and acquisition differences between UHES and ADNI from the volume scores to ensure that the patterns identified by our algorithm were impartial to those differences. We then used the grouping capabilities of our method to combine the regressed scores across hemispheres and time points associated with the same ROI to account for the interdependencies of ROI-specific measures, *i.e.*, the change between baseline and follow-up scores of the same ROI and subject was small (average correlation across all regions r=0.93) and neither HAND nor MCI expected to have an asymmetric effect on brain structure.

To the best of our knoweldge, group sparsity was the simplest mechanism with respect to logistic regression models to enforce the assumptions of bilateral and temporal consistency. Enforcing these assumptions in (non-group) sparsity constraint models, such as [Ryali et al. (2010)], required training the model on summary scores of our regional volume measures, e.g., compute the average volume of a region across time and brain hemispheres. This would have reduced the resolution of the scores available to our method making our approach less reliable in general. Alternatively, the simpler, sparsity constraint model could have been trained on the original volume scores and ignored the assumptions. However, the resulting patterns would then be difficult to interpret clinically as our method most likely would have selected only one hemisphere of an impacted structure as the other hemisphere of that region would not add significantly more information to automatically distinguish both diseases. A similar argument applied to regions selected at baseline or follow-up. The need for structured sparsity constraint beyond group sparsity, such as tree structured sparsity [Chen and Huang (2014)], would have been useful, for example, if our analysis included scores across several regions, such as the volume of the frontal lobe. However, this would have also increased the complexity of our model making it more difficult to solve, *i.e.*, the convergence of our method was unknown.

We measured the accuracy of our algorithm on this data set via five-fold cross-validation and then infered from the resulting confusion matrix the significance of the sparse, regional patterns (or image biomarkers) for differentiating MCI from HAND. We first parameterized a linear regression model based on the healthy controls of the training set and then applied it

to all samples to reduce the prevalence of factors associated with differences in data collection while aiming to preserve the differences between MCI and HAND. Based on the resulting residual score, we trained the group-sparsity constrained logistic classifier on the diseased samples. To avoid biasing the results towards a pattern with a specific number of regions, we automatically determined the number of regions to be selected via parameter exploration. On the test set, we then measured the significance of the patterns extracted by the resulting ensemble of classifiers [Rokach (2010)] in distinguishing the diseased samples from both data sets. We checked the neutrality of those patterns to differences in data collection between UHES and ADNI by computing their accuracy in correctly assigning healthy controls to the two data sets.

Our approach was to identify anatomical patterns distinguishing HAND from MCI via the following novel study design:

- We examined the differences between two diseases by fusing two independently collected data sets, each only containing cases of one disease. We ensured that the significant findings of our analysis represented differences between the two disease groups by demonstrating that the patterns were impartial, *i.e.*, not significant, to differences between control groups of the two data sets.

- Using cross-validation, we parameterized a linear regression model for data harmonization based on the healthy controls of the training set and then applied it to test sets for further analysis. Although this method has been common practice in the machine learning community, most neuroimaging studies configured those linear models on a subset of controls that were then also used for analysis [Bartsch et al. (2014); Deoni et al. (2008); van Erp et al. (2015)]. Using the same data for training and testing models, however, increased the risk for reporting overly optimistic findings [Coombes et al. (2007); Taylor et al. (2008)].

- We proposed an algorithm that determined a group sparsity constrained local minimum to a logistic classification problem. Extending [Zhang and Pohl (2015); Zhang et al. (2016)], we applied the approach to neuroimaging data for the first time. The approach simultaneously searched all regions and time points for patterns distinguishing HAND from MCI. In doing so, we identified patterns that were more accurate than those defined by individually selecting regions via more common morphometric analysis.

Materials and Methods

Samples

The Memory and Aging Center at University of California, San Francisco, provided imaging data of 15 individuals diagnosed with HIV-Associated Neurocognitive Disorder (HAND) and 22 healthy subjects. Informed consent was obtained from all participants, and the experimental protocols were approved by the institutional ethics board. All subjects were between the ages of 60 and 70. Two MRI scans were acquired of each subject, with the scans at least 10 months apart, using 3T Siemens Magnetom Trio Tim scanners. Each

subject underwent standardized comprehensive neurological and medical evaluations, including a complete neurological examination, medical history with HIV history, medications, family history, social history, laboratory tests, and a neuropsychological testing battery augmented with tests of psychomotor (grooved pegboard task) and motor speed (finger tapping) for greater sensitivity in HIV. The neuropsychological battery tapped multiple domains including memory, executive function, psychomotor speed, visuospatial and motor abilities, and attention. A proxy informant interview was completed by a trained technician using the Clinical Dementia Rating scale (CDR) [Morris (1993)].

All samples were reviewed in consensus conference meetings that included a behavioral neurologist, an HIV specialist and a neuropsychologist. This team was expert at distinguishing Alzheimers Disease from other disorders as it also reviewed cases for the UCSF Alzheimers Disease Research Center on a ragular basis. Alzheimers disease criteria and HAND criteria (guided by 2007 (Frascati) criteria [Antinori et al. (2007)]) were applied based on history, cognitive testing, neurological examination and functional testing. None of the UHES cases had clinical, neurological or cognitive testing consistent with MCI due to Alzheimer's disease.

All HAND subjects had a CDR score of 0.5. At baseline, their mean duration of HIV infection was 19.6 years, the mean CD4 count was 546, the mean CD4 nadir score was 162, 94 % of the HAND subjects were taking ART, and for 80 % of them the viral load was undetectable. The group of healthy controls was matched to the HAND group with respect to age (at baseline), gender, and 'Years between Baseline and Follow-Up', which was the difference in years between the follow-up and baseline visit across the samples of each cohort (see also the cohort specific mean and standard deviation listed in Table 1). We referred to these data as UHES.

From the ADNI project, we downloaded data of 148 Cognitive Impairment cases, 116 healthy controls scanned on 3T Siemens machines, and 96 healthy cases scanned on 3T GE scanner. Selection criteria were age range of 60 to 70, at least two MRI scanning sessions, scanners had to be acquired on a 3T GE scanner or 3T Siemens scanner. With respect to the cognitive impaired cases, we further confined the sample set to those with CDR 0.5 and that were diagnosed with MCI (Early Mild Cognitive Impairment, Mild Cognitive Impairment, or Impairment)[1]. MCI patients reported a subjective memory concern either autonomously or via an informant or clinician. They had testing impairment in memory, however, they had no significant levels of impairment in other cognitive domains, essentially preserved activities of daily living and there were no signs of dementia. HIV infection was an exclusion criteria from the ADNI study[2]. Note, while a small subset of samples of the ADNI project received amyloid testing for the diagnosis of Alzheimer's disease, all cases were diagnosed by experts comparable to those in the UHES projects. As shown in Figure 1, we then matched the three groups to the HAND cases of the UHES data set with respect to age, gender, and time between baseline and follow-up scans. The resulting matched ADNI

[1] http://adni.loni.usc.edu/study-design/background-rationale
[2] http://adni.loni.usc.edu/wp-content/uploads/2013/09/DOD-ADNI-IRB-Approved-Final-protocol-08072012.pdf

data set used for our analysis consisted of 80 MCI, 18 controls scanned on 3T Siemens, and 8 controls scanned on 3T GE machines.

Data acquisition

All MRIs were acquired on 3T scanners. For the UHES data, a Siemens Megnetom Trio Tim scanner acquired a magnetization-prepared 180 degrees radio-frequency pulses and rapid gradient-echo (MPRAGE) sequence (TR=2300ms, TI=900ms, TE=2.98 ms, flip angle=9, matrix=256×256, FOV=256mm, slice dimensions=$1.0 \times 1.0 \times 1.0$mm, 160 slices). All ADNI data was acquired using the standardized ADNI acquisition protocol[3]. 3T Siemens scanners acquired a MPRAGE sequence (TR=2300ms, TI=900ms, TE=2.98 ms, flip angle=9, matrix=256×256, FOV=256mm, slice dimensions=$1.0 \times 1.0 \times 1.2$mm, 160 slices). GE scanners acquired an Inversion Recovery-SPoiled Gradient Recalled (IR-SPGR) echo sequence (TI=400ms, flip angle=11, matrix=256×256, FOV=240mm, slice dimensions=$1.0 \times 1.0 \times 1.2$mm, 200 slices).

Image Processing

Image processing involved bias field correction via the Computational Morphometry Toolkit (CMTK) [Rohlfing (2009)]. Skull stripping was the result of majority voting [Rohlfing et al. (2004)] applied to the maps extracted by the Robust Brain Extraction (ROBEX) method [Iglesias et al. (2011)], FSL BET [Smith (2002)] and AFNI [Cox (1996)] from the native and bias field corrected T_1-weighted images. Based on the SRI24 longitudinal analysis platform [Sullivan et al. (2011a)], the follow-up, biased-corrected T_1-weighted image was non-rigidly registered to the corresponding baseline scan using the non-rigid deformation between the skull-stripped images as initialization and coarse deformation field for encoding. Afterwards, the volumes of supratentorial (svol), ventricular, cortical and cerebellar structures were extracted from the images, where the regions of interest (ROIs) were defined according to the non-rigidly aligned SRI24 atlas [Rohlfing et al. (2010)]. For each subject, this process resulted in 206 volume scores describing the corresponding brain morphometry across both time points.

Identifying Disease Specific Patterns

From those regional volume scores, we automatically extracted patterns separating HAND from MCI measuring their significance via five-fold cross validation. Each fold of the cross-validation was composed of 20% of the MCI cases, 20% of the ADNI controls scanned under GE, 20% of the ADNI controls scanned under Siemens, 20% of the UHES controls, and 20% of the HAND cases. With respect to each fold, we first harmonized the data to reduce the potential effect on pattern extraction due to differences between the ADNI and UHES data set associated with acquisition, demographics (e.g., gender or age) or healthy morphometric variation (*i.e.*, svol at baseline). We then extracted the patterns from the harmonized scores via our group-cardinality constrained, logistic regression solver [Zhang et al. (2016)]. Finally, we computed the accuracy and significance of the method and corresponding patterns. We now briefly describe our approach with the mathematical details

---

[3]see http://adni.loni.usc.edu/methods/documents/mri-protocols/ for further detail

given in the Appendix. The logistic regression software used for this analysis can be downloaded via https://dx.doi.org/10.6084/m9.figshare.3398332 or its current version via https://github.com/sibis-platform/PDLG.

We harmonized the regional volume scores by fitting a linear regression model to the baseline and follow-up measures of the controls of the training data and then computing the residual by applying the model to all samples. Measuring the quality of the residuals in harmonizing the volume scores was defined with respect to the null hypothesis that our group-sparsity constrained logistic regression model was not better than chance in correctly assigning healthy controls to the two data sets. Testing of the null hypothesis will be reviewed at the end of this section.

Next, we applied our algorithm for logistic regression problems constrained by group-cardinality [Zhang et al. (2016)] to the residual volume scores. Logistic regression identified the residual volume scores, *i.e.*, ROIs, relevant for distinguishing the two diseases while the group-cardinality encoded constraints specific to our proposed application. Specifically, we note that the change between baseline and follow-up scores of the same ROI and subject was small, *i.e.*, highly correlated (r=0.93). In addition, current medical literature is consistent in reporting absence of asymmetry in the disease effects on brain anatomy. Thus, we assumed that a residual score specific to a single time point and brain hemisphere was less informative for distinguishing the two diseases than the group of scores of that ROI across time points and hemispheres. Combining residual scores of ROI's left and right hemisphere at baseline and follow-up resulted in 52 groups. From those 52 groups, our method selected subsets, which we also refer to as patterns, as most informative with respect to distinguishing MCI from HAND. Note, this sparsity model is motivated by the specific experimental setup forcing our method to select all residual scores associated with a region (across both hemispheres and time) or not. One might choose a different grouping of scores when studying other image features or diseases, such as for disorders asymmetrically impacting brain morphometry a non-bilateral grouping would be more suitable.

To avoid biases introduced in our findings by manually fixing the size '*g*' of the pattern, *i.e.*, the only parameter to be set in our approach, we automatically determined '*g*' via parameter exploration confined to the residual scores of the MCI and HAND cases of the training set. We defined the parameter space of g by selecting values between the smallest pattern (consisting of more than one group) and 33% of the groups, *i.e.*, $g \in \{2, 3, \ldots, 17\}$. For each parameter g, we defined a group-cardinality constrained logistic regressor by computing the optimal patterns (*i.e.*, weights $\left(\breve{v}', \breve{w}' \right)$ of the regressor - see Appendix for definition). We computed the normalized accuracy (nAcc) of each regressor (or pattern) with respect to correct assignment of the disease training data to the UHES or ADNI data set, *i.e.*, we separately computed the accuracy for each cohort and then averaged their values to account for the imbalance in cohort size. On our training data, parameter exploration did not return a single optimal cardinality setting as the training accuracy of multiple regressors was 100%. We instead combined all regressors into a single ensemble of classifiers [Rokach (2010)] using the nAcc score of each regressor as the weight of the regressor in the final decision of the ensemble. In other words, the ensemble determined the label of a subject by the

weighted average across the set of regressors, all of which were unique with respect to the group-cardinality constraint. Once trained, we computed the residual scores of all the test subjects (control and disease) and recorded the ensemble's assignment of the test sample to the ADNI or UHES data set.

After testing each fold, we computed the nAcc score of our approach in correctly assigning the diseased cohorts of the test data to the ADNI or UHES data set. We also computed the p-value of the ensembles by applying the two-sided Fisher Exact Test [Fisher (1935)] to the corresponding confusion matrix. This calculation tested the null hypothesis that a classifier is not better than chance in correctly assigning samples to the two data sets. We repeated those computations for the controls. If the p-value associated with the disease cohorts were significant (p 0.001) and non-significant for the controls, then we interpreted the results generated by our approach as being informative with respect to separating HAND cases from those with MCI and impartial to difference between ADNI and UHES data.

In this scenario, we viewed patterns highly informative for separating the two cohorts as those who were part of the patterns identified by a majority of applicable regressors. As patterns extracted by regressors with larger cardinality constraints could not be sub-patterns of those extracted by regressors with smaller cardinality constraints, we defined the number of applicable regressors as those whose cardinality constraint was at last as high as the size of the pattern. Finally, we computed the frequency of appearance of each pattern, *i.e.*, the number of times a pattern was part of ones extracted by regressors divided by the number of applicable regressors, and focused further analysis on those with frequency above 50%.

Regions not selected by our approach could still be affected by either disorder. Exclusion of a region by our method indicated that the region did not contain adequate complementary information to enhance the identified pattern. Nonetheless, that hypothetical region could be part of another pattern distinguishing MCI from HAND, although that pattern might not be so robust as the one identified by our approach; alternatively, inclusion of that region might expand the size of the pattern beyond that permitted by the sparsity constraint model.

Results

We tested the proposed approach by comparing the corresponding accuracy scores and p-values to alternative approaches.

Classification Results

Table 2 summarizes the nAcc scores of applying our approach to the original volume scores (omitting harmonization) and the residual volume scores (with harmonization). The table specifies those scores with respect to distinguishing the disease cohorts and the controls from both data sets. Furthermore, the table shows the results of training and testing the regression model with three different data configurations: one just on the baseline score ($T = 1$), one just on the follow-up data ($T = 1$), and one using the data from both time points ($T = 2$) (as proposed in the previous section). Scores in bold refer to the experiments in which p-values were significantly different (p 0.001) from randomly assigning samples to each data set.

When omitting data harmonization, the ensemble was more accurate in distinguishing the controls than the disease from both data sets. The opposite was true when applying our approach to the (harmonized) residual scores. Now, the scores in distinguishing the control groups from both data sets were not statistically significant, while separating cases of HAND from MCI was significant for those experiments including measurements from follow-up visits. Distinguishing HAND from MCI relying solely on the follow-up measures resulted in an nAcc score of 81.5% (vs. 64.2% when only using baseline measures). Combining baseline and follow-up measurements resulted in the most accurate separation (88.1%) between the two cohorts while the accuracy in distinguishing the control groups (61.4%, p-value=0.1) was within the range defined by the two cross-sectional experiments.

Pattern Analysis

We now further review the patterns obtained by the most accurate ensemble, which was based on residual scores of the baseline and follow-up visits of all 102 regions. In our experiment, the ensembles of classifiers identified 78 unique patterns, whose frequencies were plotted in Figure 2. Of those, the frequency of 8 patterns were above 50%. The 8 patterns comprised 8 bilateral regions (see also Figure 3): medial orbitofrontal cortex, cerebellum VIII, parahippocampal gyrus, precuneus, superior temporal pole, precentral gyrus, cerebellum VIIb, and inferior temporal gyrus.

If we reduced the threshold to patterns with frequency above 20%, *i.e.*, the frequency before the curve in Figure 2 flattens out, then 9 more patterns were included in the analysis. However, these extra patterns only contained two more regions: fusiform cortex and postcentral gyrus. These two regions appeared only once across those patterns and across all patterns no more than 25 times, which was much fewer than the appearance of any of the other 8 regions (46 to 71 times). We concluded that the first 8 patterns with the associated 8 regions were most relevant with respect to distinguishing HAND from MCI.

Table 3 lists the 8 selected regions in the order of their number of appearances across all 78 patterns, the regions associated with each pattern, and the accuracy of each pattern (last two rows) and region (last two columns) in distinguishing the cohorts. For each region and pattern, we computed the accuracy and p-value (two-sided Fischer Exact Test) via five-fold cross-validation of the logistic regression (without any cardinality constraints) applied to the corresponding residual scores.

The p-value in distinguishing the controls of the two data sets was not significant (*i.e.*, $p > 0.001$) for any pattern or region. By contrast, all patterns except Pattern 7 were significant in distinguishing the disease groups, whereas the disease specific classification accuracy for regions was never higher than 68.1%, the lowest accuracy across all patterns was 68.3% for Pattern 7 and the highest was 90.8% for Pattern 3. The 3 patterns with the highest accuracy (Pattern 3,5,6) were also the only ones that included the 3 regions most frequently identified: the medial orbitofrontal cortex, cerebellum VIII, and parahippocampal gyrus. Interestingly, our experiments relying on just each region individually reported non-significant p-values with respect to distinguishing the two disease cohorts. The three regions that resulted in significant p-values were the cerebellum VIIb, precuneus, and inferior temporal gyrus. The remaining 2 of the 8 regions identified by the top 8 patterns, the superior temporal pole and

the precentral gyrus, neither had significant p-values nor emerged as part any of the top three patterns.

Morphometric Analysis

To compare our findings on the 8 regions with more conventional morphometric volumetric analysis, we parameterized the linear regression model of Section Data Harmonization based on the baseline volume scores of all controls and computed the residuals for all volume scores (including disease and follow-up visit). For each of the 8 regions, we computed the slopes between follow-up and baseline residual scores. Table 4 lists the corresponding p-values when applying the Mann-Whitney U test [Mann and Whitney (1947)] to those slopes. Entries in bold were significantly different (p   0.001) between the corresponding cohorts. For those regions, Figure 4 showed the box plots of the slopes based on the residual volume scores for all cohorts.

The trajectory for none of the regions was significantly different between MCI and controls. However, the trajectories of 3 regions were significantly different for HAND vs. MCI (right cerebellum VIIb, right cerebellum VIII, left precentral) and HAND vs. controls (left precuneus, right precentral, right cerebellum VIIb). For these regions, subjects with HAND aged significantly faster than the corresponding other cohorts (see also Figure 4). For the other regions, the residual scores regressed much faster for the HAND cases than the other groups. However, the change was not significant. The cerebellum VIIb was the only region that significantly distinguished HAND from MCI and HAND from controls. It was also the only region that overlapped with the significant findings of the region specific classification of the previous section.

Discussion

Without data harmonization, the classifier encoded differences in data collection between the ADNI and the UHES data set as it was more accurate in distinguishing controls than disease groups. This was not the case for the classifier based on the (harmonized) residual scores, which identified patterns distinguishing HAND from MCI on a significant level while not being significantly affected by differences in data collection (for experiment based on baseline and follow-up $p = 0.1$). Identifying such a bias in our patterns would require reapplying our method to a larger data set. Based on our experience from Zhang et al. (2016) and the results of Table 2, training our approach on a larger data set generally increases its accuracy in distinguishing MCI from HAND, *i.e.*, results in improved patterns. These improved patterns might also bring the accuracy associated with the control groups closer to 50%, which would then confirm again that our analysis is unbiased towards differences in data collection between the two sets.

When relying solely on follow-up scores, the accuracy in distinguishing HAND from MCI was higher than just based on the ones for the baseline scans reflecting the understanding that the underlying patterns were more distinct as both degeneration progressed. Combining baseline and follow-up measures resulted in the most accurate separation, which was significantly better ($p < 0.001$) than solely relying on the baseline scores according to the Delongs test [DeLong et al. (1988)]. These findings underline the strength of trajectories

extracted from the longitudinal analysis over cross-sectional volume scores in identifying differences between HAND and MCI. Our experiments also showed that each of the 8 highest ranked patterns was also more informative in differentiating cases of HAND from MCI than each of the 8 regions associated with those patterns as the disease specific classification accuracy of the patterns was always higher than the accuracy scores of any single region. In other words, the significance of the identified image biomarkers, *i.e.*, patterns, for distinguishing MCI from HAND is much higher than differences based solely on the individual regions within those patterns.

Of the 3 most frequently identified regions (medial orbitofrontal cortex, cerebellum VIII, and parahippocampal), the medial orbitofrontal cortex, which is involved in the cognitive processing of decision-making, had been found to be abnormally small in MCI and Alzheimer's patients 4 years before the occurrence of any cognitive symptoms [Tondelli et al. (2012)]. Nothing had been reported on the role of MCI on the cerebellum VIII or more generally the cerebellum, which is known to subserve motor tasks and likely be involved in selective cognitive functions commonly associated with the frontal lobe functions [Schmahmann and Pandya (2008)]. However, in comparison to healthy controls, higher atrophy rates in the cerebellum have been published for HIV+ [Klunder et al. (2008); Elsheikh et al. (2010); Chang et al. (2011); Sullivan et al. (2011b)]. Finally, heightened atrophy in the parahippocampal gyrus, which contributes to memory encoding and retrieval, has been reported in HIV+ [Maki et al. (2009); Wilson et al. (2015)] and MCI cohort [Pantel et al. (2003)].

Next, we reviewed the literature for the three regions associated with significant p-values, *i.e.*, cerebellum VIIb, precuneus, and inferior temporal gyrus. Whereas volume deficits of the cerebellum VIIb had not been associated with the functional deficits of HAND or MCI, we noted above the higher atrophy rates in the cerebellum of HIV+ subjects. The precuneus, however, which is associated with episodic memory and consciousness, is one of the few structures known to atrophy faster in MCI subjects than healthy controls [Tondelli et al. (2012); Pihlajamaki et al. (2009)]. The inferior temporal gyrus has also been reported to have higher atrophy rates in MCI and HIV exhibiting cognitive decline [Küper et al. (2011); McDonald et al. (2012)]. This region is associated with high level visual processing, such as the shape recognition [Martin et al. (1996)]. Finally, the regions that neither had significant p-values nor emerged as part any of the top three patterns, *i.e.*, the superior temporal pole and the precentral gyrus, were also the only regions where neither MCI nor HIV had been associated significant volume deficits.

From our morphometric analysis on those 8 regions, we concluded that the patterns found by our logistic regression approach were mostly influenced by HAND as significant differences in regional trajectories could only be found in group comparisons including this disease. With respect to at least one hemisphere of 4 regions, HAND had significantly faster tissue loss than the other cohorts (see also Figure 4). These findings were not only consistent with the medical literature but also with the diminished function observed in HAND samples. The cerebellum VIIb was the only region where the significant findings of the morphometric analysis agreed with those of the logistic classifier. We thus concluded that common morphometric analysis might not directly translate to decision making on an individual

subject level. To investigate this possibility further, we computed the p-value of all 102 regions. We then trained a logistic regressor on those 14 regions, whose slope was significantly different between MCI and HAND at least for one hemisphere. The 14 regions were the middle frontal gyrus, rolandic operculum, anterior cingulate gyrus, calcarine sulcus, lingual gyrus, middle occipital, fusiform gyrus, inferior parietal lobule, angular gyrus, transverse temporal gyri, crus I of cerebellar hemisphere, crus II of cerebellar hemisphere, lobule VIIB of cerebellar hemisphere, and lobule VIII of cerebellar hemisphere. Five-fold cross-validation resulted in a nAcc scores of 73.8% for distinguishing HAND from MCI, whereas for the control group it was 56.1%. In comparison, when we trained the classifier on the 8 regions selected by the 8 top patterns, then the accuracy increased to 94.8% for distinguishing the disease cohorts while the accuracy dropped to 47.5% for the control groups. These results underline the importance of selecting patterns of regions (as it was done by our group cardinality constrained logistic regressor) instead of individual regions when separating HAND from MCI.

Conclusion

We proposed a group sparsity constrained logistic classifier to identify morphometric patterns separating HAND from MCI in patients between the ages of 60 and 70. Confining classifiers to sparsity constrained solutions has been popular in medical image analysis as they reduce the risk of overfitting of the classifier on small sample sizes, which is a common characteristic for neuroimaging studies. These classifiers approximated the sparse solution by changing the underlying regularization from counting non-zero weights via the l_0-'norm' to analyzing the weights' absolute values by applying the l_1-norm or, in case of group sparsity constrained problems, the l_2-norm. These approximations have brought in question the resulting solutions with respect to the original problem. We therefore suggested herein an algorithm that directly solves the original, sparse problem.

We applied our proposed classifier to a data set containing 15 cases of HAND provided by the UHES study and 80 age matched MCI cases downloaded from the ADNI project. Using cross-validation, we trained a linear model on healthy controls of both studies to harmonize the volumetric scores for demographic and acquisition differences across the two data sets while preserving disease specific pattern distinguishing HAND from MCI. We identified patterns of various sizes by training our proposed group sparsity constrained logistic classifier on samples of both diseases. During testing, we then showed that data harmonization was successful as the accuracy of those patterns was non-significant in assigning healthy controls to the correct data set. However, the patterns were significant in correctly assigning disease cases to HAND or MCI with the longitudinal study design being more accurate than the cross-sectional analysis.

The most frequently identified patterns were composed of a total of 8 regions. Our findings were largely in line with the literature, which reported on some of these regions as having measured volume deficits in HIV subjects, being significant for characterizing the early onset of MCI, or predicting subjects with MCI converting to Alzheimer's' disease. Furthermore, our morphmetric group analysis revealed that HAND had significantly faster tissue loss than MCI or controls in at least one hemisphere in half of those regions, *i.e.*, the

right cerebellum VIIb, right cerebellum VIII, and left precentral with respect to MCI and the left precuneus, right precentral, right cerebellum VIIb with respect to controls.

Using morphometric group analysis to individually select regions resulted in a pattern that was larger in size and less informative than the most frequently identified patterns by our proposed approach. We therefore conclude that simultaneously analyzing all brain regions and time points for disease specific patterns, as done by our group sparsity constrained logistic classification, enhances the ability to distinguish HAND from cognitive impairment found in the non-HIV infected population, such as the MCI cohort.

Acknowledgments

Appendix: Model and Numerical Solution to Pattern Extraction

We extracted patterns separating HAND from MCI by first applying logistic regression constrained by group-cardinality [Zhang et al. (2016)] to harmonized volume scores. We harmonized the regional volume scores via linear regression. For each ROI 'i', the linear regression model was defined by the following factors for subject 's' at time point t: the volume score $v_{i,t}^s$, the baseline svol score v_{svol}^s, u_{ADNI-G}^s and u_{ADNI-S}^s encoding the acquiring scanner (see Table 5 for details), and the age at scanning $u_{age_t}^s$. More specifically, the following linear regression model was fitted to the control group:

$$v_{i,t}^s \sim \alpha_{i,0} + \alpha_{i,1} u_{age_t}^s + \alpha_{i,2} u_{svol}^s + \alpha_{i,3} u_{ADNI-G}^s + \alpha_{i,4} u_{ADNI-S}^s \quad (1)$$

Based on this fitting, the residuals were computed for all subjects (including those with MCI or HAND)

$$r_{i,t}^s := v_{i,t}^s - (\alpha_{i,0} + \alpha_{i,1} u_{age_t}^s + \alpha_{i,2} u_{svol_t}^s + \alpha_{i,3} u_{ADNI-G}^s + \alpha_{i,4} u_{ADNI-S}^s). \quad (2)$$

Based on the residual volume scores $r_{i,t}^s$, our group-cardinality constrained logistic regression model aims to automatically identify the regional patterns distinguishing the two diagnosis. To define the model, we assumed (without loss of generality) training of the model was based on $N = 76$ samples of the MCI and HAND cases. $\{b_1,\ldots, b_N\}$ encoded the assignment of the N samples to the two data sets, where $b_s \in \{-1, +1\}$ was $+1$ if subject 's' was a part of UHES, *i.e.*, HAND, and -1 otherwise. The input of the model were the residual volume scores

$$r^s := \left[r_{1,1}^s \ldots \ldots r_{R,T}^s \right]^\top$$

of $R = 102$ ROIs extracted over $T = 2$ timepoints. Combining r^s with b_s, we defined the label weighted feature vectors $a^s := b_s \cdot r^s$. The model then encoded the influence of each element in a^s in correctly classifying the subject 's' via the weight vector 'w'.

To determine the optimal $\hat{w}$ with respect to separating the two cohorts while being consistent across time and hemisphere, we coupled with each ROI $\tilde{i}$ a group $G_{\tilde{i}}$, which consisted of the indexes associated with the ROI's left and right hemisphere at baseline and follow-up. This resulted in $M = 52$ groupings $\{G_1,\ldots, G_M\}$. We then used those groupings to define $w_{G_{\tilde{i}}}$, the subvector of w indexed by $G_{\tilde{i}}$, and $\|w_{G_{\tilde{i}}}\|_2$, the l_2-norm of that grouping. Based on those definitions, the group cardinality of w was then simply the number of nonzero components of

$$\tilde{w} := (\|w_{G_1}\|_2, \ldots, \|w_{G_M}\|_2)^\top, \quad (3)$$

which was formally determined via the l_0-"norm" $\|\tilde{w}\|_0$. For our model to identify meaningful patterns, we further assumed that no more than $g \in \mathbb{N}$ groups with $g \quad M$ differentiate cases of HAND from MCI, *i.e.*,

$$\|\tilde{w}\|_0 \leq g. \quad (4)$$

Having defined the group cardinality constraint, we assumed that the optimal $\hat{w}$ was the result of minimizing the *average logistic loss* function, which, for label weight scalar v and w, was defined as

$$l_{\text{avg}}(v, w) := \frac{1}{N} \sum_{s=1}^{N} \theta\left(w^\top a^s + v \cdot b_s\right) \quad (5)$$

with

$$\theta(x) := \log\left(1 + \exp\left(-x\right)\right) \quad (6)$$

being the logistic function. In other words, our model was fully described by the following minimization problem

$$(\hat{v}, \hat{w}) := \underset{v \in \mathbb{R}, w \in \mathbb{R}^{R \cdot T}}{arg\min} l_{\mathrm{avg}}(v, w) \quad \text{subject to (s.t.)} \quad \|\tilde{w}\|_0 \leq g. \tag{7}$$

Alternatively to choose ROIs across time and hemispheres, one could have selected individual residual volume scores of w by replacing $\|\tilde{w}\|_0$ with $\|w\|_0$. The result would have been a *sparse logistic regression* model, which, however ignored our early assumptions of the dependency of regional scores across time and hemispheres.

We defined a possible pattern for distinguishing the two diseases as the set of ROIs whose corresponding values in $\hat{w}$ were non-zero. In other words, we were only interested in ROIs selected by our model rather than the absolute values of $\hat{w}$, which, according to [Haufe et al. (2014); Sabuncu (2014)], should not be used to study the impact of diseases on brain anatomy. To find a numerical solution for $\hat{w}$, we computed a local minimum to Eq. (7) according to [Zhang et al. (2016)] by decoupling the minimization of $l_{\mathrm{avg}}(\cdot, \cdot)$ from the sparsity constraint

$$\mathscr{X} := \{ w \in \mathbb{R}^{R \cdot T} : \tilde{w} := (\|w_{G_1}\|_2, \ldots, \|w_{G_M}\|_2)^\top \text{ and } \|\tilde{w}\|_0 \leq g \}.$$

Specifically, we introduced the auxiliary variable $\hat{y} \in \mathbb{R}^{R \cdot T}$ as an unconstrained approximation to the sparse solution $\tilde{w} \in \mathscr{X}$ by rewriting Eq. (7) as

$$(\hat{v}, \hat{y}, \hat{w}) := \underset{v \in \mathbb{R}, y \in \mathbb{R}^{R \cdot T}, w \in \mathscr{X}}{arg\min} l_{\mathrm{avg}}(v, y) \quad \text{s.t.} \quad w - y = 0. \tag{8}$$

We iteratively solved Eq. (8) via penalty decomposition (PD), *i.e.*, we approximated $(\hat{v}, \hat{y}, \hat{w})$ via the local minimum of the following non-convex problem defined with respect to the penalty parameter $\varrho > 0$:

$$\left(\breve{v}, \breve{y}, \breve{w} \right) \leftarrow \underset{v \in \mathbb{R}, y \in \mathbb{R}^{R \cdot T}, w \in \mathscr{X}}{arg\min} l_{\mathrm{avg}}(v, y) + \frac{\varrho}{2} \|w - y\|_2^2. \tag{9}$$

Initializing ϱ at 0.1 (according to [Lu and Zhang (2013)]), PD iterated between determining a local minimum of the above equation via Block Coordinate Descent (BCD) and increased ϱ by a factor of $\sqrt{10}$ (set according to [Lu and Zhang (2013)]) until convergence, *i.e.*,

$$\| \breve{w} - \breve{y} \|_F^2 \leq 10^{-3}, \tag{10}$$

which was the termination criterion for PD suggested in [Lu and Zhang (2013)].

BCD alternated between minimizing Eq. (9) with fixed w and by fixing v and y. When w was set to w', finding the minimum with respect to v and y turned into solving the smooth and convex problem

$$(v', y') \leftarrow \underset{v \in \mathbb{R}, y \in \mathbb{R}^{R \cdot T}}{arg\min} \left\{ l_{\text{avg}}(v, y) + \frac{\varrho}{2} \| w' - y \|_2^2 \right\} \quad (11)$$

via gradient descent. In turn, minimizing the objective function just with respect to w, *i.e.*,

$$w' \leftarrow \underset{w \in \mathscr{X}}{arg\min} \| w - y' \|_2^2, \quad (12)$$

could now be solved in closed form. To do so, we sorted $\| y'_{G_i} \|_2$ in descending order and recorded the order of the indexes as $\{ j_1, \dots j_M \}$, *i.e.*,

$$\| y'_{G_{j_1}} \|_2 \geq \| y'_{G_{j_2}} \|_2 \geq \dots \geq \| y'_{G_{j_M}} \|_2 \geq 0. \quad (13)$$

From Eq. (12), w' had to be equal to y' for the first 'g' groups and otherwise zero, *i.e.*, $w'_{G_{j_i}} := y'_{G_{j_i}}$ for i r and otherwise $w'_{G_{j_i}} := 0$.[4] In theory, this solution was not unique as it could happen that $\| y'_{G_{j_r}} \|_2 = \| y'_{G_{j_{r+1}}} \|_2$ so that an alternative solution was defined by $w'_{G_{j_i}} := y'_{G_{j_i}}$ for i r, $w'_{G_{j_{r+1}}} := y'_{G_{j_{r+1}}}$ and otherwise $w'_{G_{j_i}} := 0$. However, we had not experienced this issue in practice.

Proving the convergence of PD to a local minimum of Eq. (7) was non-trivial (see also [Zhang et al. (2016)]) as Eq. (9) required a search within the group-sparsity constrained space $\mathscr{X}$, which was non-convex and non-continuous. However, one could generalize the findings of [Lu and Zhang (2013)] to group sparsity to gain a better understanding of the convergence properties of the approach. Specifically, at each iteration of PD, the BCD approach generated a sequence of w'. Assuming that the sequence converged to an accumulation point $\breve{w}$, then $\breve{w}$ was also a local minimum of Eq. (11). Furthermore, PD produced a sequence of $\breve{w}'$. If the sequence converged to an accumulation point $\breve{w}'$ and $\| \breve{w}' \|_0 = g$ then $\breve{w}'$ was a local minimum of Eq. (7). In our experiments, these conditions where always fulfilled so that convergence of our implementation was not an issue and proving convergence for the more general case of $\| \breve{w}' \|_0 \leq g$ was not needed. Also note, our algorithm can only determine a local solution, we showed in Zhang et al. (2016) that this local solution is superior to the global solution of an existing method that relaxes the group sparsity constraint.

---

[4]The proof was provided by the supplemental information of [Zhang et al. (2016)].

References

Ances BM, Hammoud DA. Neuroimaging of HIV-associated neurocognitive disorders (HAND). Current Opinion in HIV and AIDS. 2014; 9(6):545–551. [PubMed: 25250553]

Ances BM, Ortega M, Vaida F, Heaps J, Paul R. Independent effects of HIV, aging, and HAART on brain volumetric measures. Journal of Acquired Immune Deficiency Syndromes. 2012; 59(5):469–477. [PubMed: 22269799]

Antinori A, Arendt G, Becker J, Brew B, Byrd D, Cherner M, Clifford D, Cinque P, Epstein L, Goodkin K, Gisslen M, Grant I, Heaton R, Joseph J, Marder K, Marra C, McArthur J, Nunn M, Price R, Pulliam L, Robertson K, Sacktor N, Valcour V, Wojna V. Updated research nosology for hiv-associated neurocognitive disorders. Neurology. 2007; 69(18):1789–1799. [PubMed: 17914061]

Bartsch H, Thompson WK, Jernigan TL, Dale AM. A web-portal for interactive data exploration, visualization, and hypothesis testing. Frontiers in Neuroinformatics. 2014; 8

Brew BJ, Crowe S, Landay A, Cysique LA, Guillemin G. Neurodegeneration and ageing in the HAART era. Journal of Neuroimmune Pharmacology. 2009; 4(2):163–174. [PubMed: 19067177]

Candès EJ, Romberg J, Tao T. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. IEEE Transactions on Information Theory. 2006; 52(2):489–509.

Cardenas VA, Meyerhoff DJ, Studholme C, Kornak J, Rothlind J, Lampiris H, Neuhaus J, Grant RM, Chao LL, Truran D, Weiner MW. Evidence for ongoing brain injury in human immunodeficiency virus-positive patients treated with antiretroviral therapy. Journal of Neurovirology. 2009; 15(4):324–333. [PubMed: 19499454]

Carroll MK, Cecchi GA, Rish I, Garg R, Rao AR. Prediction and interpretation of distributed neural activity with sparse models. NeuroImage. 2009; 44(1):112–122. [PubMed: 18793733]

Chang L, Andres M, Sadino J, Jiang C, Nakama H, Miller E, Ernst T. Impact of apolipoprotein E $\mu4$ and HIV on cognition and brain atrophy: antagonistic pleiotropy and premature brain aging. NeuroImage. 2011; 58(4):1017–1027. [PubMed: 21803164]

Chang L, Holt JL, Yakupov R, Jiang CS, Ernst T. Lower cognitive reserve in the aging human immunodeficiency virus-infected brain. Neurobiology of Aging. 2013; 34(4):1240–1253. [PubMed: 23158761]

Chen C, Huang J. The benefit of tree sparsity in accelerated MRI. Medical image analysis. 2014; 18(6):834–842. [PubMed: 24380657]

Cohen RA, Seider TR, Navia B. HIV effects on age-associated neurocognitive dysfunction: premature cognitive aging or neurodegenerative disease? Alzheimer's Research & Therapy. 2015; 7(37):1–10.

Coombes KR, Wang J, Baggerly KA. Microarrays: retracing steps. Nature Medicine. 2007; 13(11):1276–1277.

Cox RW. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Computers and Biomedical research. 1996; 29(3):162–173. [PubMed: 8812068]

Cysique LA, Brew BJ. The effects of HIV and aging on brain functions: proposing a research framework and update on last 3 years' findings. Current Opinion in HIV and AIDS. 2014; 9(4):355–364. [PubMed: 24871088]

Cysique LA, Moffat K, Moore DM, Lane TA, Davies N, Carr A, Brew BJ, Rae C. HIV, vascular and aging injuries in the brain of clinically stable HIV-infected adults: A ^{1}H MRS study. PloS one. 2013; 8(4):e61738. [PubMed: 23620788]

DeLong E, DeLong D, Clarke-Pearson D. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. Sep; 1988 44(3):837–845. [PubMed: 3203132]

Deoni SC, Williams SC, Jezzard P, Suckling J, Murphy DG, Jones DK. Standardized structural magnetic resonance imaging in multicentre studies using quantitative T1 and T2 imaging at 1.5T. NeuroImage. 2008; 40(2):662–671. [PubMed: 18221894]

Elsheikh BH, Maher WE, Kissel JT. Cerebellar atrophy associated with human immunodeficiency virus infection. Archives of Neurology. 2010; 67(5):634–635. [PubMed: 20457966]

Ettenhofer ML, Hinkin CH, Castellon SA, Durvasula R, Ullman J, Lam M, Myers H, Wright MJ, Foley J. Aging, neurocognition, and medication adherence in HIV infection. The American Journal of Geriatric Psychiatry. 2009; 17(4):281–290. [PubMed: 19307857]

Fisher RA. The logic of inductive inference. Journal of the Royal Society. 1935; 98:38–54.Series A

Focke NK, Helms G, Kaspar S, Diederich C, Tóth V, Dechent P, Mohr A, Paulus W. Multi-site voxel-based morphometry—not quite there yet. NeuroImage. 2011; 56(3):1164–1170. [PubMed: 21324367]

Friedman J, Hastie T, Tibshirani R. A note on the group Lasso and a sparse group Lasso. 2010 arXiv preprint arXiv:1001.0736.

Han X, Jovicich J, Salat D, van der Kouwe A, Quinn B, Czanner S, Busa E, Pacheco J, Albert M, Killiany R, Maguire P, Rosas D, Makris N, Dale A, Dickerson B, Fischl B. Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. NeuroImage. 2006; 32(1):180–194. [PubMed: 16651008]

Haufe S, Meinecke F, Görgen K, Dähne S, Haynes JD, Blankertz B, Bießmann F. On the interpretation of weight vectors of linear models in multivariate neuroimaging. NeuroImage. 2014; 87:96–110. [PubMed: 24239590]

Heaton RK, Franklin DR, Ellis RJ, McCutchan JA, Letendre SL, LeBlanc S, Corkran SH, Duarte NA, Clifford DB, Woods SP, Collier AC, Marra CM, Morgello S, Mindt MR, Taylor MJ, Marcotte TD, Atkinson JH, Wolfson T, Gelman BB, McArthur JC, Simpson DM, Abramson I, Gamst A, Fennema-Notestine C, Jernigan TL, Wong J, Grant I. HIV-associated neurocognitive disorders before and during the era of combination antiretroviral therapy: differences in rates, nature, and predictors. Journal of Neurovirology. 2011; 17(1):3–16. [PubMed: 21174240]

Heaton RK, Marcotte TD, Mindt MR, Sadek J, Moore DJ, Bentley H, McCutchan JA, Reicks C, Grant I. The impact of HIV-associated neuropsychological impairment on everyday functioning. Journal of the International Neuropsychological Society. 2004; 10(03):317–331. [PubMed: 15147590]

Holt JL, Kraft-Terry SD, Chang L. Neuroimaging studies of the aging HIV-1-infected brain. Journal of Neurovirology. 2012; 18(4):291–302. [PubMed: 22653528]

Iglesias JE, Liu CY, Thompson PM, Tu Z. Robust brain extraction across datasets and comparison with publicly available methods. IEEE Transactions on Medical Imaging. 2011; 30(9):1617–1634. [PubMed: 21880566]

Jahanshad N, Valcour VG, Nir TM, Kohannim O, Busovaca E, Nicolas K, Thompson PM. Disrupted brain networks in the aging HIV+ population. Brain connectivity. 2012; 2(6):335–344. [PubMed: 23240599]

Jovicich J, Marizzoni M, Sala-Llonch R, Bosch B, Bartrés-Faz D, Arnold J, Benninghoff J, Wiltfang J, Roccatagliata L, Nobili F, et al. Brain morphometry reproducibility in multi-center 3T MRI studies: a comparison of cross-sectional and longitudinal segmentations. NeuroImage. 2013; 83:472–484. [PubMed: 23668971]

Kirk JB, Goetz MB. Human immunodeficiency virus in an aging population, a complication of success. Journal of the American Geriatrics Society. 2009; 57(11):2129–2138. [PubMed: 19793157]

Klunder AD, Chiang MC, Dutton RA, Lee SE, Toga AW, Lopez OL, Aizenstein HJ, Becker JT, Thompson PM. Mapping cerebellar degeneration in HIV/AIDS. Neuroreport. 2008; 19(17):1655. [PubMed: 18806691]

Küper M, Rabe K, Esser S, Gizewski E, Husstedt I, Maschke M, Obermann M. Structural gray and white matter changes in patients with HIV. Journal of Neurology. 2011; 258(6):1066–1075. [PubMed: 21207051]

Li S, Yin H, Fang L. Group-sparse representation with dictionary learning for medical image denoising and fusion. IEEE Transactions on Biomedical Engineering. 2012; 59(12):3450–3459. [PubMed: 22968202]

Liu A, Press DZ. Does HIV age your brain? Neurology. 2013; 80(13):1178–1179. [PubMed: 23446676]

Liu M, Zhang D, Shen D. Ensemble sparse classification of Alzheimer's disease. NeuroImage. 2012; 60(2):1106–1116. [PubMed: 22270352]

Lu Z, Zhang Y. Sparse approximation via penalty decomposition methods. SIAM Journal on Optimization. 2013; 23(4):2448–2478.

Lv J, Jiang X, Li X, Zhu D, Chen H, Zhang T, Zhang S, Hu X, Han J, Huang H, Zhang J, Guo L, Liu T. Sparse representation of whole-brain fMRI signals for identification of functional networks. Medical Image Analysis. 2015; 20(1):112–134. [PubMed: 25476415]

Ma S, Huang J. Penalized feature selection and classification in bioinformatics. Briefings in Bioinformatics. 2008; 9(5):392–403. [PubMed: 18562478]

Maki P, Cohen M, Weber K, Little D, Fornelli D, Rubin L, Perschler P, Gould F, Martin E. Impairments in memory and hippocampal function in HIV-positive vs HIV-negative women a preliminary study. Neurology. 2009; 72(19):1661–1668. [PubMed: 19433739]

Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. The Annals of Mathematical Statistics. 1947; 18(1):50–60.

Martin A, Wiggs CL, Ungerleider LG, Haxby JV. Neural correlates of category-specific knowledge. Nature. 1996; 379(6566):649–652. [PubMed: 8628399]

McArthur JC, Steiner J, Sacktor N, Nath A. Human immunodeficiency virus-associated neurocognitive disorders: Mind the gap. Annals of Neurology. 2010; 67(6):699–714. [PubMed: 20517932]

McDonald CR, Gharapetian L, McEvoy LK, Fennema-Notestine C, Hagler DJ, Holland D, Dale AM. Relationship between regional atrophy rates and cognitive decline in mild cognitive impairment. Neurobiology of Aging. 2012; 33(2):242–253. [PubMed: 20471718]

McMurtray A, Nakamoto B, Shikuma C, Valcour V. Cortical atrophy and white matter hyperintensities in HIV: the Hawaii aging with HIV cohort study. Journal of Stroke and Cerebrovascular Diseases. 2008; 17(4):212–217. [PubMed: 18589342]

Meier L, Van De Geer S, ühlmann P. The group Lasso for logistic regression. Journal of the Royal Society. 2008; 70:53–71.Series B

Morris JC. The clinical dementia rating (cdr): current version and scoring rules. Neurology. 1993; 43(11):2412–2414.

Ng, B., Vahdat, A., Hamarneh, G., Abugharbieh, R. Machine Learning in Medical Imaging. 2010. Generalized sparse classifiers for decoding cognitive states in fMRI; p. 108-115.Vol. 6357 of Lecture Notes in Computer Science

Nir TM, Jahanshad N, Busovaca E, Wendelken L, Nicolas K, Thompson PM, Valcour VG. Mapping white matter integrity in elderly people with HIV. Human Brain Mapping. 2014; 35(3):975–992. [PubMed: 23362139]

Pantel J, Kratz B, Essig M, Schröder J. Parahippocampal volume deficits in subjects with aging-associated cognitive decline. American Journal of Psychiatry. 2003; 160(2):379–382. [PubMed: 12562591]

Pfefferbaum A, Rogosa DA, Rosenbloom MJ, Chu W, Sassoon SA, Kemper CA, Deresinski S, Rohlfing T, Zahr NM, Sullivan EV. Accelerated aging of selective brain structures in human immunodeficiency virus infection: a controlled, longitudinal magnetic resonance imaging study. Neurobiology of Aging. 2014; 35(7):1755–1768. [PubMed: 24508219]

Pfefferbaum A, Rohlfing T, Pohl KM, Lane B, Chu W, Kwon D, Nichols BN, Brown SA, Tapert SF, Cummins K, Thompson WK, Brumback T, Meloy MJ, Jernigan TL, Dale A, Colrain IM, Baker FC, Prouty D, De Bellis MD, Voyvodic JT, Clark DB, Luna B, Chung T, Nagel BJ, Sullivan EV. Adolescent development of cortical and white matter structure in the NCANDA sample: role of sex, ethnicity, puberty, and alcohol drinking. Cerebral Cortex. 2015 In press.

Pfefferbaum A, Rohlfing T, Rosenbloom MJ, Chu W, Colrain IM, Sullivan EV. Variation in longitudinal trajectories of regional brain volumes of healthy men and women (ages 10 to 85years) measured with atlas-based parcellation of MRI. NeuroImage. 2013; 65:176–193. [PubMed: 23063452]

Pihlajamaki M, Jauhiainen AM, Soininen H. Structural and functional MRI in mild cognitive impairment. Current Alzheimer Research. 2009; 6(2):179–185. [PubMed: 19355853]

Rao, A., Lee, Y., Gass, A., Monsch, A. Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE. IEEE; 2011. Classification of Alzheimer's disease from structural MRI using sparse logistic regression with optional spatial regularization; p. 4499-4502.

Rohlfing, T. User guide to the computational morphometry toolkit. The Insight Journal. 2009. http://hdl.handle.net/10380/3140

Rohlfing T, Russakoff DB, Maurer CR Jr. Performance-based classifier combination in atlas-based image segmentation using Expectation-Maximization parameter estimation. Medical Imaging, IEEE Transactions on. 2004; 23(8):983–994.

Rohlfing T, Zahr NM, Sullivan EV, Pfefferbaum A. The SRI24 multichannel atlas of normal adult human brain structure. Human Brain Mapping. 2010; 31(5):798–819. [PubMed: 20017133]

Rokach L. Ensemble-based classifiers. Artificial Intelligence Review. 2010; 33(1–2):1–39.

Rosa MJ, Portugal L, Hahn T, Fallgatter AJ, Garrido MI, Shawe-Taylor J, Mourao-Miranda J. Sparse network-based models for patient classification using fMRI. NeuroImage. 2015; 105:493–506. [PubMed: 25463459]

Ryali S, Supekar K, Abrams DA, Menon V. Sparse logistic regression for whole-brain classification of fMRI data. NeuroImage. 2010; 51(2):752–764. [PubMed: 20188193]

Sabuncu, M. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014. 2014. A universal and efficient method to compute maps from image-based prediction models; p. 353-360.Vol. 8675 of Lecture Notes in Computer Science

Schmahmann JD, Pandya DN. Disconnection syndromes of basal ganglia, thalamus, and cerebro-cerebellar systems. Cortex. 2008; 44(8):1037–1066. special Issue on 'Brain Hodology - Revisiting disconnection approaches to disorders of cognitive function'. URL http://www.sciencedirect.com/science/article/pii/S0010945208001196. [PubMed: 18614161]

Simmons A, Westman E, Muehlboeck S, Mecocci P, Vellas B, Tsolaki M, Kłoszewska I, Wahlund L-O, Soininen H, Lovestone S, Evans A, Spenger C. The AddNeuroMed framework for multicentre MRI assessment of Alzheimer's disease: experience from the first 24 months. International Journal of Geriatric Psychiatry. 2011; 26(1):75–82. [PubMed: 21157852]

Smith SM. Fast robust automated brain extraction. Human Brain Mapping. 2002; 17(3):143–155. [PubMed: 12391568]

Stonnington CM, Tan G, Klöppel S, Chu C, Draganski B, Jack CR, Chen K, Ashburner J, Frackowiak RS. Interpreting scan data acquired from multiple scanners: a study with Alzheimer's disease. NeuroImage. 2008; 39(3):1180–1185. [PubMed: 18032068]

Sullivan EV, Pfefferbaum A. Diffusion tensor imaging in normal aging and neuropsychiatric disorders. European Journal of Radiology. 2003; 45(3):244–255. [PubMed: 12595109]

Sullivan EV, Pfefferbaum A, Rohlfing T, Baker FC, Padilla ML, Colrain IM. Developmental change in regional brain structure over 7 months in early adolescence: comparison of approaches for longitudinal atlas-based parcellation. NeuroImage. 2011a; 57(1):214–224. [PubMed: 21511039]

Sullivan EV, Rosenbloom MJ, Rohlfing T, Kemper CA, Deresinski S, Pfefferbaum A. Pontocerebellar contribution to postural instability and psychomotor slowing in HIV infection without dementia. Brain Imaging and Behavior. 2011b; 5(1):12–24. [PubMed: 20872291]

Taylor JM, Ankerst DP, Andridge RR. Validation of biomarker-based risk prediction models. Clinical Cancer Research. 2008; 14(19):5977–5983. [PubMed: 18829476]

Thomas JB, Brier MR, Snyder AZ, Vaida FF, Ances BM. Pathways to neurodegeneration effects of HIV and aging on resting-state functional connectivity. Neurology. 2013; 80(13):1186–1193. [PubMed: 23446675]

Tondelli M, Wilcock GK, Nichelli P, Jager CAD, Jenkinson M, Zamboni G. Structural MRI changes detectable up to ten years before clinical Alzheimer's disease. Neurobiology of Aging. 2012; 33(4):825.e25–36.

Towgood KJ, Pitkanen M, Kulasegaram R, Fradera A, Kumar A, Soni S, Sibtain NA, Reed L, Bradbeer C, Barker GJ, Kopelman MD. Mapping the brain in younger and older asymptomatic HIV-1 men: Frontal volume changes in the absence of other cortical or diffusion tensor abnormalities. Cortex. 2012; 48(2):230–241. [PubMed: 21481856]

Valcour V, Shikuma C, Shiramizu B, Watters M, Poff P, Selnes O, Holck P, Grove J, Sacktor N. Higher frequency of dementia in older HIV-1 individuals the Hawaii aging with HIV-1 cohort. Neurology. 2004; 63(5):822–827. [PubMed: 15365130]

van Erp TGM, Hibar DP, Rasmussen JM, Glahn DC, Pearlson GD, Andreassen OA, Agartz I, Westlye LT, Haukvik UK, Dale AM, Melle I, Hartberg CB, Gruber O, Kraemer B, Zilles D, Donohoe G,

Kelly S, McDonald C, Morris DW, Cannon DM, Corvin A, Machielsen MWJ, Koenders L, de Haan L, Veltman DJ, Satterthwaite TD, Wolf DH, Gur RC, Gur RE, Potkin SG, Mathalon DH, Mueller BA, Preda A, Macciardi F, Ehrlich S, Walton E, Hass J, Calhoun VD, Bockholt HJ, Sponheim SR, Shoemaker JM, van Haren NEM, Pol HEH, Ophoff RA, Kahn RS, Roiz-Santianez R, Crespo-Facorro B, Wang L, Alpert KI, Jonsson EG, Dimitrova R, Bois C, Whalley HC, McIntosh AM, Lawrie SM, Hashimoto R, Thompson PM, Turner JA. Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the enigma consortium. Molecular Psychiatry. 2015 In press.

Vounou M, Janousova E, Wolz R, Stein JL, Thompson PM, Rueckert D, Montana G. Sparse reduced-rank regression detects genetic associations with voxel-wise longitudinal phenotypes in Alzheimer's disease. NeuroImage. 2012; 60(1):700–716. [PubMed: 22209813]

Westman E, Simmons A, Muehlboeck JS, Mecocci P, Vellas B, Tsolaki M, Kłoszewska I, Soininen H, Weiner MW, Lovestone S, Spenger C, Wahlund LO. AddNeuroMed and ADNI: similar patterns of Alzheimer's atrophy and automated MRI classification accuracy in europe and north america. NeuroImage. 2011; 58(3):818–828. [PubMed: 21763442]

Wilson TW, Heinrichs-Graham E, Becker KM, Aloi J, Robertson KR, Sandkovsky U, White ML, O'Neill J, Knott NL, Fox HS, Swindells S. Multimodal neuroimaging evidence of alterations in cortical structure and function in HIV-infected older adults. Human Brain Mapping. 2015; 36(3): 897–910. [PubMed: 25376125]

Wu F, Yuan Y, Zhuang Y. Heterogeneous feature selection by group Lasso with logistic regression. Proceedings of the International Conference on Multimedia. 2010:983–986.

Xiang S, Yuan L, Fan W, Wang Y, Thompson PM, Ye J. Bi-level multi-source learning for heterogeneous block-wise missing data. NeuroImage. 2014; 102:192–206. [PubMed: 23988272]

Yamashita O, Sato M-a, Yoshioka T, Tong F, Kamitani Y. Sparse estimation automatically selects voxels relevant for the decoding of fMRI activity patterns. NeuroImage. 2008; 42(4):1414–1429. [PubMed: 18598768]

Ye J, Farnum M, Yang E, Verbeeck R, Lobanov V, Raghavan N, Novak G, DiBernardo A, Narayan VA. Sparse learning and stability selection for predicting mci to ad conversion using baseline adni data. BMC Neurology. 2012; 12(1):1–12. [PubMed: 22289169]

Yuan L, Wang Y, Thompson PM, Narayan VA, Ye J. Multi-source feature learning for joint analysis of incomplete multiple heterogeneous neuroimaging data. Neuroimage. 2012; 61(3):622–632. [PubMed: 22498655]

Zhang D, Shen D. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. NeuroImage. 2012; 59(2):895–907. [PubMed: 21992749]

Zhang S, Zhan Y, Metaxas DN. Deformable segmentation via sparse representation and dictionary learning. Medical Image Analysis. 2012; 16(7):1385–1396. [PubMed: 22959839]

Zhang Y, Kwon D, Pohl KM. Computing group cardinality constraint solutions for logistic regression problems. Medical Image Analysis. 2016 In press.

Zhang Y, Pohl KM. Solving logistic regression with group cardinality constraints for time series analysis. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. 2015:459–466.Volume 9351 of the series Lecture Notes in Computer Science

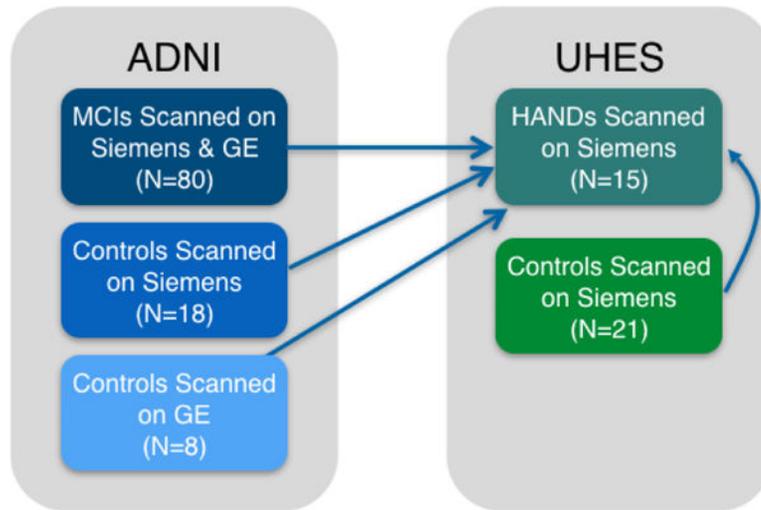**Figure 1.**
Matching demographic of cohorts **of the ADNI and UHES data set** to the HAND **samples**.
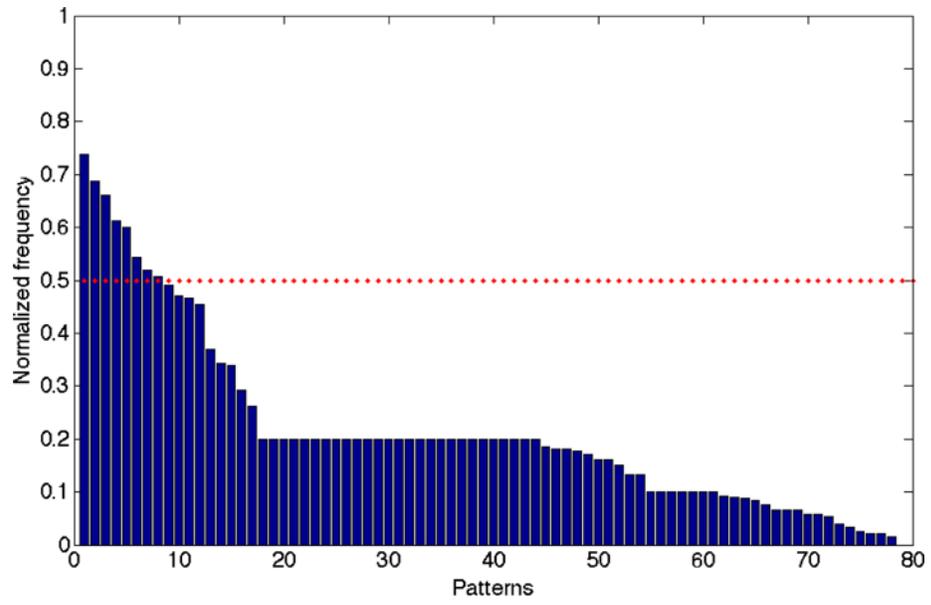
Figure 2.
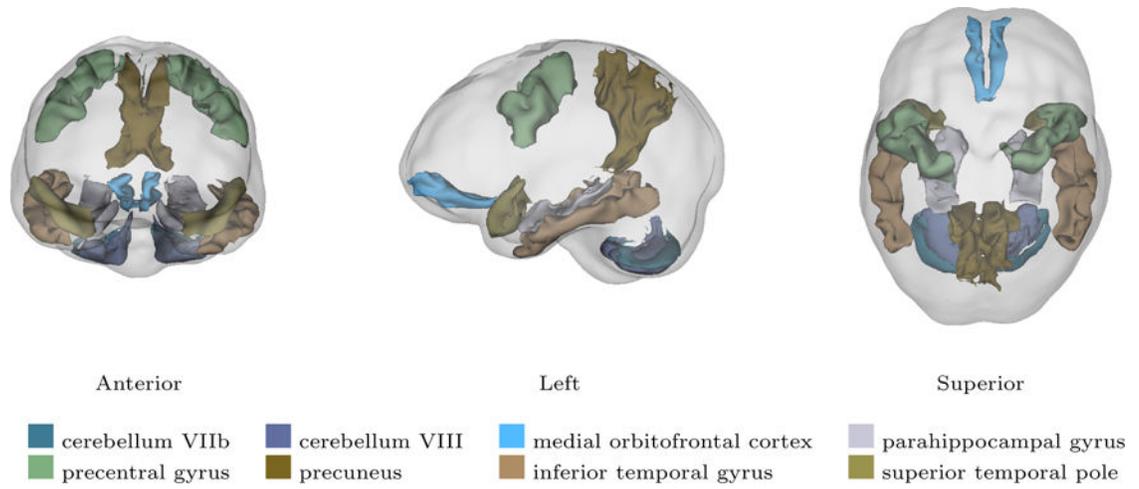Frequency of patterns across all regressors of the ensemble

Figure 3.
3D models of the selected regions.
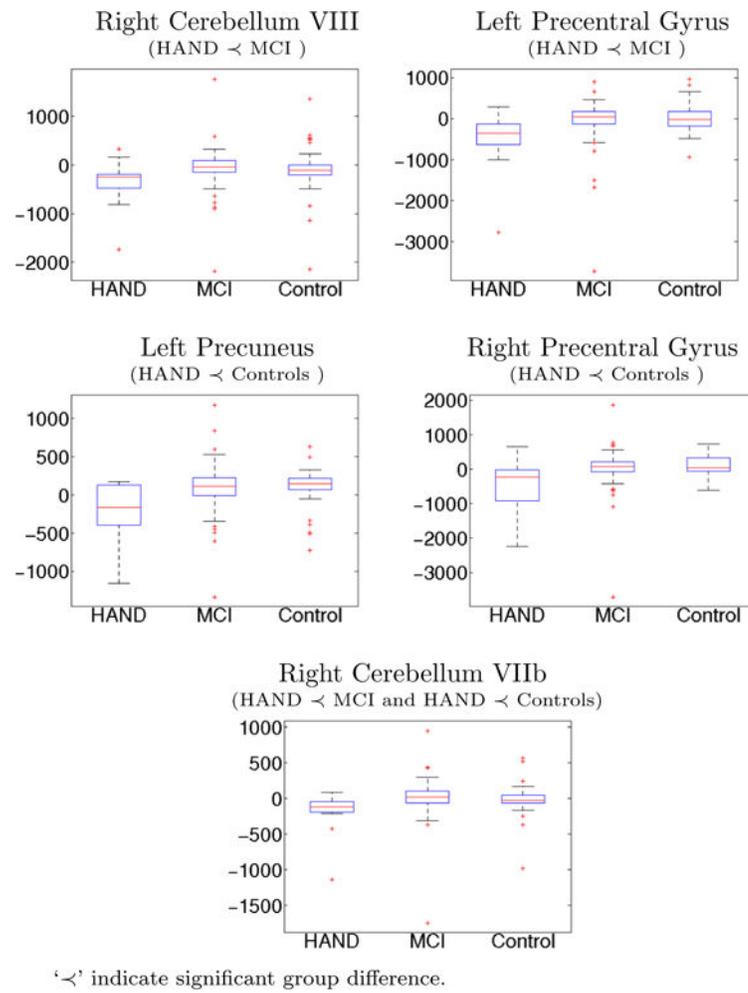
Figure 4.
The slopes of the residual scores for regions with significant p-value.

Table 1

Demographic information of the two data sets.

| | UHES | | MCI | ADNI | |
| | HAND | Control | | Control | |
				Siemens	GE
Number of subjects	15	21	80	18	8
Gender (M/F)	14/1	19/2	70/10	15/3	7/1
Age (mean±SD)	63.8 ± 3.3	65.1 ± 3.2	65.1 ± 3.2	65.2 ± 3.2	65.6 ± 3.2
Years difference between Baseline Follow-up and (mean ± SD)	1.9 ± 0.9	2.1 ± 0.9	2.1 ± 0.9	2.3 ± 1.2	1.9 ± 0.7

Table 2

Accuracy scores in % for different implementations.

Implementation	Set	Baseline	Follow-Up	Baseline + Follow-Up
Original Volume Scores (no harmonization)	Disease	**73.8%**	**70.8%**	**77.7%**
	Control	**82.8%**	**77.6%**	**79.9%**
Residual Scores (with harmonization)	Disease	64.2%	**81.5%**	**88.1%**
	Control	61.8%	60.9%	61.4%

Scores in **bold** refer to experiments in which p-values were significantly different (p 0.001) from randomly assigning samples to each data set.

Table 3

Patterns with their relative frequency being above 50%.

Regions	% of Appearn.	Patterns								nAcc(%)	
		1	2	3	4	5	6	7	8	Dis.	Ctr.
medial orbitofrontal cortex	8.8	x			x	x	x			64.8	54.8
cerebellum 8	83.8		x		x	x	x	x	x	61.5	51.4
parahippocampal gyrus	75.0	x		x		x	x	x		52.1	50.0
precuneus	75.0		x	x		x			x	**65.4**	50.5
superior temporal pole	66.3				x				x	58.8	48.5
precentral gyrus	62.5						x			63.5	55.2
cerebellum 7b	62.5			x		x		x		**68.1**	53.8
inferior temporal gyrus	57.5					x				**65.4**	57.1
nAcc for Disease groups (%)		**70.2**	**70.8**	**90.8**	**76.3**	**87.5**	**88.1**	**68.3**	**70.2**		
nAcc for Control groups(%)		54.8	51.8	44.6	53.3	44.1	52.3	53.8	54.2		

Scores in **bold** referred to the experiments whose p-values were significant (p 0.001).

Table 4

p-values of Mann-Whitney U tests based on volume change (slope).

Regions	HAND vs. MCI		HAND vs. Control		MCI vs. Control	
	Left	Right	Left	Right	Left	Right
medial orbitofrontal cortex	0.084	0.003	0.038	0.005	0.770	0.813
cerebellum 8	0.082	**<0.001**	0.285	0.005	0.246	0.108
parahippocampal gyrus	0.034	0.988	0.028	0.178	0.524	0.031
precuneus	0.003	0.043	**<0.001**	0.010	0.417	0.609
superior temporal pole	0.295	0.010	0.118	0.002	0.381	0.386
precentral gyrus	**0.001**	0.002	0.005	**0.001**	0.537	0.755
cerebellum 7b	0.015	**<0.001**	0.005	**0.001**	0.254	0.066
inferior temporal gyrus	0.003	0.010	0.003	0.020	0.863	0.616

Values in **bold** were significant (p 0.001).

Table 5

Encoding for scanners.

Scanner	u^s_{ADNI-G}	u^s_{ADNI-S}
ADNI Siemens	0	1
ADNI GE	1	0
UHES Siemens	0	0