A brain signature highly predictive of future progression to Alzheimer's dementia

Christian Dansereau[a,b,*], Angela Tam[c], AmanPreet Badhwar[a], Sebastian Urchs[c], Pierre Orban[a,e,f], Pedro Rosa-Neto[d], Pierre Bellec[a,b,*], for the Alzheimer's Disease Neuroimaging Initiative[**]

[a] *Centre de Recherche de l'Institut Universitaire de Gériatrie de Montréal, Montréal, CA*
[b] *Département d'Informatique et de recherche opérationnelle, Université de Montréal, Montréal,CA*
[c] *Integrated Program in Neuroscience, McGill University, Montréal,CA*
[d] *Douglas Mental Health institute, McGill University, Montréal,CA*
[e] *Centre de Recherche de l'Institut Universitaire en Santé Mentale de Montréal, Montréal, CA*
[f] *Département de Psychiatrie, Université de Montréal, Montréal, CA*

Abstract

Alzheimer's disease develops slowly over years or even decades before the emergenceapparition of clinical symptoms and, eventually, dementia onset. Preliminary research Many works have aimed at finding biomarkers able to accurately predict future progression to dementia in individuals with mild cognitive impairment. Unfortunately, patients diagnosed with Alzheimer's dementia represent a highly heterogeneous group from the standpoint of the brain pathophysiological perspectivey. Accurate prediction of progression to dementia more than one year before onset has therefore proved very challenging. In this work, we propose a new machine learning technique that identifies a subgroup of patients for which clinical predictions can be made with high precision, i.e. the vast majority of selected individuals will eventually progress to dementia. We demonstrate here that it is indeed possible to train a model to predict future progression to dementia withreach high specificity (97%) and precision (90%), when predicting future progression to dementia up to three years before onset, in the cohort assembled by the Alzheimer's disease neuroimaging initiative (ADNI). The model only achieved moderate sensitivity (47%) because it was designed to target a specific brain signature mixing spatial patterns of atrophy and functional dysconnectivity. This multimodal, highly predictive brain signature was extracted from magnetic resonance images only, yet it was systematically

[*] Corresponding authors: christiandansereau@gmail.com, pierre.bellec@criugm.qc.ca

[**] Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: `http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf`

accompanied by deposition of beta amyloid plaques, a hallmark of Alzheimer's disease. Our results represent a marked improvement on the current state of the art in terms of specificity (about 10% increase) and precision (about 16% increase). The signature was in addition identified on patients suffering from dementia first, demonstrating that typical patterns of neurodegeneration are already present in prodromal individuals. Our approach provides a feasible method to select individuals with very high risk of progressing to dementia several years before the onset. We believe this technology has a lot of potential to enrich the recruitment in clinical trials, and help demonstrate, or invalidate, the efficacy of new interventions. Our The method we used is relatively simple, buildsing on well established machine learning tools, and may prove useful in the future for a wide range of other applications, particularly where the targets for predictions are noisy or heterogeneous, as is often the case in medicine.

---

1. Introduction

Alzheimer's disease (AD) is the most common age-related neurodegenerative disorder. The typical progression of late-onset, sporadic AD comprises a lengthy preclinical stage, a prodromal stage of mild cognitive impairment (MCI), and a final stage of dementia. Usually, by the time patients experience the dementia phase, severe and irreversible neurodegeneration has already occurred. In order to be effective, therapies should likely be initiated at earlier stages of the disease, when some markers of the disease are apparent but the symptoms have yet not appeared. For this reason, many works have aimed at finding biomarkers that can predict future progression to AD dementia at the prodromal or even preclinical stages (Rathore et al., 2017; Orban et al., 2017). Accurate prediction beyond two years has however proven to be challenging, likely due to the considerable pathophysiological heterogeneity underlying existing clinical diagnosis (Rathore et al., 2017). We propose here to address the heterogeneity issue by identifying a subset of individuals with MCI who share a common brain signature predictive of oncoming AD dementia with high precision.

A clinical diagnosis of Alzheimer's dementia is primarily established on the basis of symptoms. To qualify for dementia, these symptoms need to interfere with a patient's ability to function in daily activities. Dementia is considered to be probably due to AD when the symptoms appear gradually and the most prominent deficits fall either into an amnestic (i.e. memory, the most common) or nonamnestic category, i.e. language, visual or executive. There also needs to be no dominant symptoms suggestive of another type of neuropathology such as Lewy bodies, fronto-temporal atrophy or vascular abnormalities (McKhann et al., 2011).

The actual cause of dementia, AD or otherwise, can currently only be confirmed by post mortem pathophysiological examination. The hallmarks of AD are the accumulation of amyloid beta plaques and tau protein neurofibrillary tangles in the brain, as well as marked atrophy of the medial temporal lobe. The analysis of Beach et al. (2012) revealed an important mismatch between

clinical and histopathological diagnoses: sensitivity ranged from 71% to 87% and specificity ranged from 44% to 71%, depending on the level of confidence in the clinical and pathophysiological examination. In particular, 30% of patients diagnosed with AD dementia in that study had no or very minimal signs of AD pathology in their brain. In addition to such incorrect diagnoses, comorbidity of neurodegenerative diseases was highly prevalent, that is a co-occurrence of two or more disorders including AD, cerebrovascular disease, Lewy body disease, or frontotemporal degeneration Rabinovici et al. (2017); Jellinger et al. (2014). Biomarkers of AD can also be observed in 10% to 30% of cognitively normal (CN) individuals, as well as 40% of patients diagnosed with non-AD dementia (Beach et al., 2012). Finally, plaques and tangles are general markers of brain injury that are not unique to AD, e.g. they are seen in patients with brain traumatic injuries (Marklund et al., 2009). Distinct pathways are likely involved in AD, and subtypes of AD pathophysiology may emerge in the future (Au et al., 2015). In summary, the clinical labels of neurodegeneration currently used are often incorrect (wrong underlying disease), incomplete (missing several interacting diseases) or unspecific (pooling together different pathways with overlapping biomarkers).

To better diagnose AD in vivo, many imaging techniques have been developed to track the propagation of key markers, both across brain regions and over time. Both amyloid beta and tau can be imaged in vivo using Positron Emission Tomography (PET) (Fodero-Tavoletti et al., 2011; **?**). Structural magnetic resonance imaging (MRI) provides a non-invasive measure of temporal lobe atrophy, as well as decreases in cortical thickness throughout the brain (Lerch et al., 2005). Large imaging samples such as the one collected by the Alzheimers Disease Neuroimaging Initiative (ADNI) have established that amyloid beta and tau starts accumulating years, possibly decades, before the onset of clinical symptoms, and that atrophy also typically precedes the onset of clinical symptoms (McConathy and Sheline, 2015). Imaging biomarkers are increasingly used to complement neuropsychological testing to diagnose AD (Dubois et al., 2007). In recent years, a great amount of work has been devoted to the identification of novel or more sensitive imaging based biomarkers of AD and MCI using machine learning techniques (Rathore et al., 2017). The current state of the art on predictive models using the ADNI dataset reached 95% accuracy (precision of 96%, specificity of 95% and sensitivity of 92%) to classify AD vs cognitively normal (CN) (Fan et al., 2008b; Zhu et al., 2014; Xu et al., 2015; Zu et al., 2016) and 80% accuracy (precision of 80%, specificity of 75% and sensitivity 85%) to identify patients with MCI who will progress to AD dementia in the next three years (Toussaint et al., 2012; Moradi et al., 2015; Zheng et al., 2015; Cheng et al., 2015a,b; Korolev et al., 2016), using mainly anatomical, FDG-PET or amyloid-PET measures. The prediction accuracy for progression to dementia however plummeted after 1.5 years, reaching only 75% accuracy over a 18-36 months time window (Korolev et al., 2016; Arbabshirani et al., 2017).

These accuracy scores however need to be properly interpreted. Korolev et al. (2016), in particular, took great care of reporting separately the specificity (76%, proportion of stable MCI being correctly classified), sensitivity (83%,

proportion of progressor MCI being correctly classified), and precision (80%), i.e. the proportion of actual progressors amongst individuals classified as such. Precision is a key metric for enrichment in a clinical trial, as it dictates how many patients will decline in the absence of treatment. For a given sensitivity and specificity, the precision does depend on the baseline ratio between stable and progressor MCI in the sample. Working on a scenario of 30% MCI progressors in the cohort (which matches well actual rates seen in clinical populations), Korolev et al. (2016) concluded to an expected precision of 60.2%. There is therefore ample margin for improvements in terms of prognostic precision of future progression to AD dementia within 3 years. We note that, because there are more MCI stables than progressors at baseline, an increase in specificity has higher leverage on precision than sensitivity. For example, working again on a scenario of 30% MCI progressors in the cohort, with an equal sensitivity and specificity at 80%, the precision is 63%. An increase of 10% sensitivity will only increase precision by 3%, while an increase of 10% specificity will boost precision by 14%.

We hypothesized that the precision of imaging-based diagnosis of AD in past studies was severely limited by the pathophysiological heterogeneity of clinical cohorts. In this work, we proposed a new machine learning technique that aims at identifying a subgroup of patients for whom clinical predictions can be made with high precision. We specifically trained a model to sacrifice sensitivity for high specificity and precision (Figure 1A). The behaviour of the proposed method is illustrated by a simple simulation (Figure 1B). The task was to classify two classes using a separation line, represented by blue dots for controls and red dots for patients. The distribution of both red and blue subjects was heterogeneous, in the sense that each distribution was a mixture of several Gaussian classes. Some of these classes were clearly separable, yet others were not, with blue and red points closely overlapping (maybe because of incorrect, incomplete or unspecific diagnoses). When a standard classifier is applied on that data, it identifies a separation line making a tradeoff in sensitivity and specificity across all examples (see Figure 1B, second column). By perturbing the data, it is possible to identify the easy cases, i.e. the data point that can be correctly and reliably classified: more opaque points are associated with more reliable predictions and clearly identify the two well-separated classes at the top in Figure 1B, third column. A separate model is then trained to identify the easy cases red points (see Figure 1B, fourth columns). The resulting prediction of red labels has limited sensitivity, as the problematic cases are not being detected at all, but it has near perfect specificity and precision. Note that the example found in Figure 1B was actually computed with our proposed method. To evaluate how this method managed to extract AD biomarkers, we first examined the classic problem of predicting clinical diagnosis in a cohort including CN participants and patients with AD dementia. We then used the model trained on the CN vs AD classification problem to make predictions on the subjects with MCI. Our hypothesis was that a brain signature of AD dementia would already be present at the prodromal stage, and predictive of future progression to dementia. To test this hypothesis, we evaluated whether the MCI patients flagged as AD would

develop AD dementia within three years.

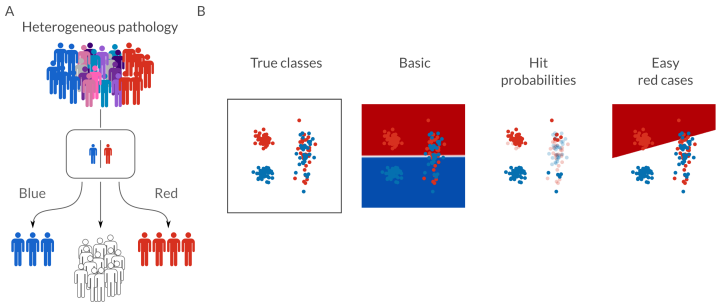

Figure 1: Panel A show the identification of easy cases for each class, Panel B prediction of clinical labels in a two-class problem, in the presence of heterogeneous labels in a subset of the data. First column show the initial classification problem with the distribution of the two classes. The second column show a basic classifier decision hyperplane. Third column show the subject that have been flagged as high hit probability in hard color and the low hit probabilities with some transparency. Fourth column show the final decision hyperplane of the red subject with the HPS signature.

2. Results

Multimodal imaging markers

We extracted multimodal measures of brain organization, that could be used for automated AD diagnosis. The measures were derived from the baseline MRI scans of the Alzheimer's Disease Neuroimaging Initiative 2 (ADNI2) cohort, which included anatomo-functional imaging for CN subjects (N=49) as well as patients suffering from AD dementia (N=24) (available sample size after quality control on 10/2016). We decided to include a range of different measures as a basis for diagnosis, which have previously been shown to be sensitive markers of AD dementia. These included gray matter (GM) thickness (Querbes et al., 2009; Eskildsen et al., 2013), GM volume of various brain structures (Karas et al., 2004), as well as seed-based fMRI connectivity maps generated for 20 intrinsic connectivity brain networks (Bellec et al., 2015).

Substantial inter-individual variations were observed in the brain distribution of imaging measures. For example, some subjects showed higher- or lower-than average volumetric measures across extensive brain territories, such as the right medial occipital cortex in subject 1 (lower) and subject 73 (higher), see Figure 2A. We investigated whether such patterns could be found systematically in a subgroup of subjects. For this purpose, we quantified the similarity of GM volume maps between any given pair of subjects using a Pearson correlation coefficient (Figure 2B). A cluster analysis revealed the presence of three subgroups of subjects with homogeneous GM volume maps. These subgroups were apparent as diagonal squares on the inter-subject similarity matrix with high similarity values, Figure 2B. These squares outline all subject-to-subject

similarities within a specific subgroup. By contrast, low similarity values were observed in elements outside of these squares, which corresponded to pairs of subjects falling into different subgroups. A subtype template was generated for each subgroup by averaging maps of individuals within that subgroup, Figure 2B). In particular, subtypes 2 and 3 of GM volumetric maps reproduced the pattern in the occipital cortex observed in subjects 1 and 73, respectively. The separation between clusters was not clear-cut in matrix 2B, suggesting a continuum rather than discrete subtypes. We thus extracted a continuous measure (Pearson's correlation) of similarity, called subtype weight, between each individual map and each subtype map, Figure 2D). The subtyping procedure outlined above was applied independently for each type of measure (volumetric, cortical thickness, rs-fMRI) and each brain network (for rs-fMRI). We concluded by visual inspection to the presence of at least three subtypes for each modality/network, which we thus selected as a common number of subtypes across all modalities/networks for subsequent analyses.

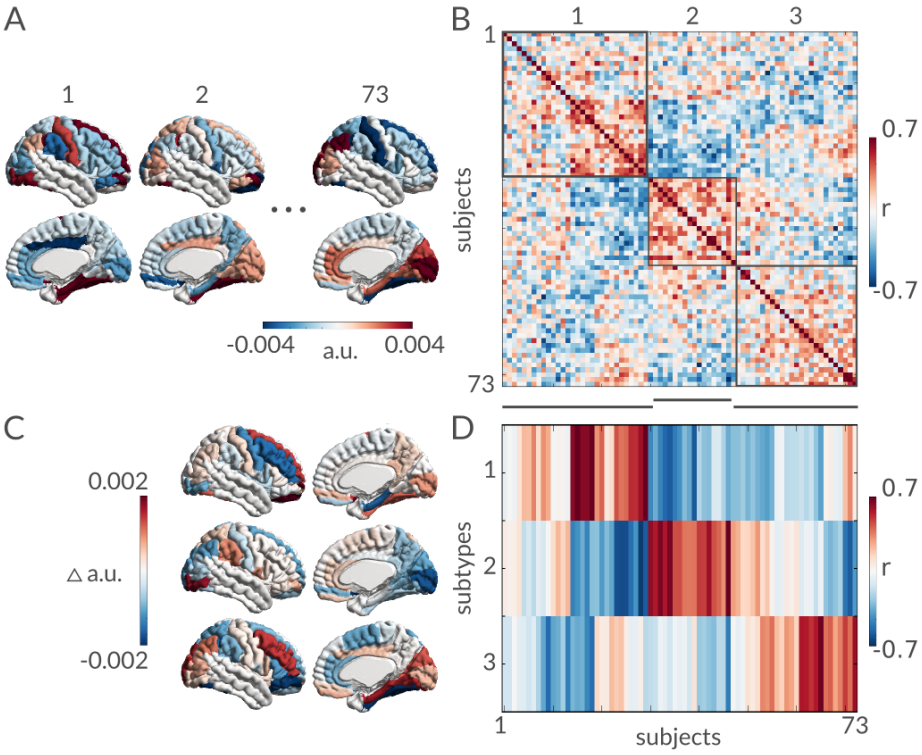

Figure 2: Demeaned gray matter volume measures of the right hemisphere. Panel A shows individual maps and the correlation of every subject with all other subjects in Panel B. Panel C shows the subtypes templates representing subgroups in the dataset. Panel D shows the association of each individual map in A with the each subtype template in C.

Prediction of AD

We established a baseline performance for automatic classification of CN vs AD subjects using a well established machine learning model, i.e. a linear support vector machine model (SVM) (Cortes and Vapnik, 1995). The model reached 70% precision (specificity 86%, sensitivity 67%) using tenfold cross-validation and multimodal (fMRI + sMRI) subtype weights, Figure 3. The performance of the method trained on only fMRI 38% precision (specificity 47% and sensitivity 67%) and sMRI data only had a very close performance, with 67% precision (specificity 84%, sensitivity 67%). Note that, during cross-validation, the training of the model included both the generation of subtypes and the optimization of the SVM parameters.

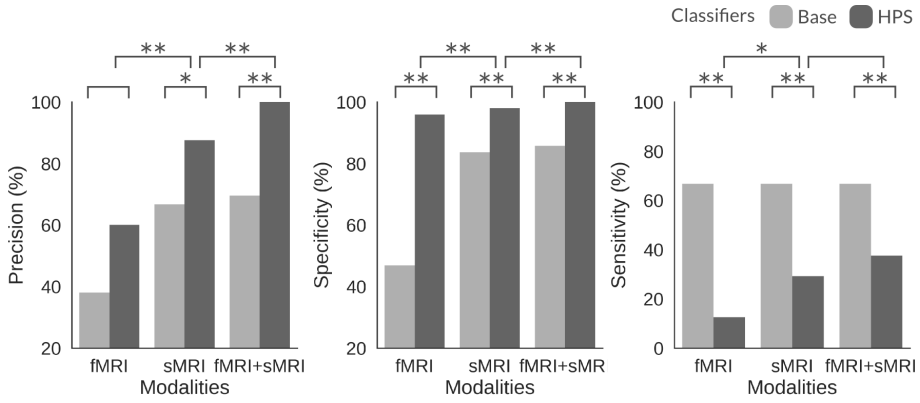

Figure 3: Figure shows the precision, specificity and sensitivity of the three modalities (fMRI, sMRI and fMRI+sMRI) at each stage (Base: basic classifier and HPS: highly predictive signature). Significant differences are shown with $*$ for $p < 0.05$ and $**$ for $p < 0.001$).

Identifying easy cases

As we outlined in the introduction, the core idea of our approach was to identify a subset of subjects for which clinical labels are easy to predict, such as the points on the left in Figure 1A. To identify these "easy cases", we randomly perturbed the input data of the SVM model many times through subsampling, and assessed the hit probability for any given subject to be properly classified. We found that 68% of individuals had a perfect (100%) hit probability, with a small subset of subjects (18%) exhibiting less reliable predictions (hit-probability $< 90\%$), Supplementary material S1). We defined the "easy cases" as the subgroup of individuals reaching perfect hit probability.

Predicting easy cases

The next step of the method was to train a logistic regression (Fan et al., 2008a) to predict the AD "easy cases" Figure 6B, analogous to the rightmost column of Figure 1B. The full multi-stage process of subtype extraction, hit

probability estimation and logistic regression was cross-validated using a ten-fold scheme in order to generate the performance of the prediction of AD "easy cases". A perfect 100% precision (specificity 100%, sensitivity 36%) was reached for AD "easy cases", using multimodal structural and functional features. A significant improvement (in precision and specificity $p < 0.001$) of the HPS compared to the performance of the method trained on only fMRI data reached a precision of 60% (specificity 96%, sensitivity 13%) and sMRI data reached a precision of 88% (specificity 98%, sensitivity 29%), see see Figure 3. Compared to the reference SVM model, with multimodal features, the precision of our proposed HPS model was improved by a wide margin (30%, $p < 0.001$), as well as the specificity (15%, $p < 0.001$), at the cost of a marked loss in sensitivity (30%, $p < 0.001$). See Supplementary material Table S2 for a list of the performance of each model.

Highly predictive brain signature

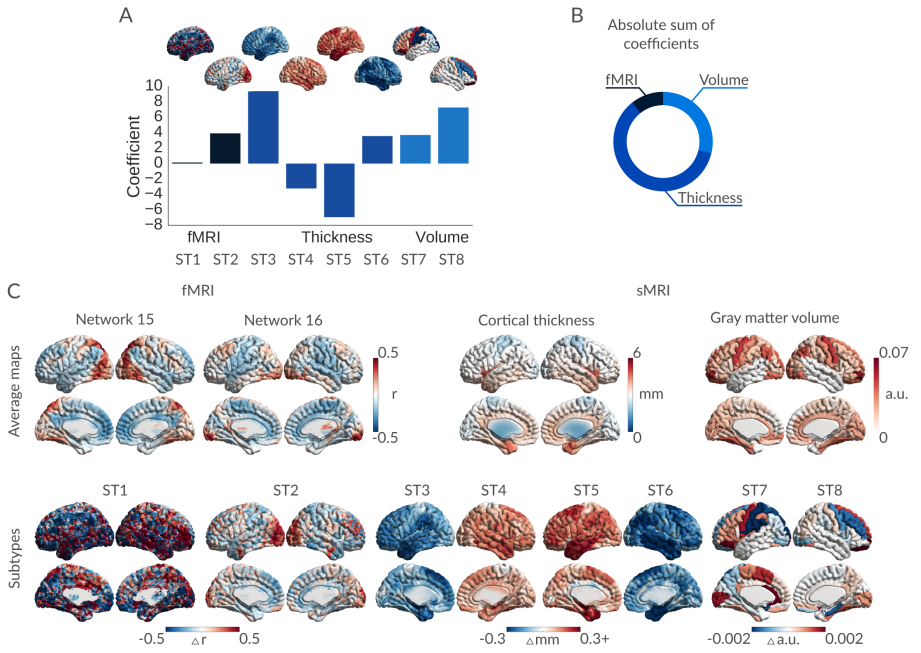

Figure 4: Panel A shows the contribution of each modality to the decision, the ratios are computed by the sum of the absolute coefficient for each modality. Panel B shows the coefficients of the high-confidence prediction model for each subtype map. Panel C shows, on top, the average maps for each modality and on the bottom the subtype maps used for the high-confidence prediction.

The logistic regression model used to predict AD easy cases is based on a set of coefficients, which give more or less weight to a particular subtype and modality. As such, the individuals flagged as AD easy cases can be seen as sharing a brain HPS, composed of combination of subtype maps. The logistic

model may in theory ignore a subtype or an entire modality, by setting the corresponding weights to zero. In practice, we found that the HPS relied on all three types of measures (functional connectivity, cortical thickness, and gray matter volume), Figure 4A. To rank the contribution of each modality in the decision process, we computed the absolute sum of the coefficients for each measure, relative to the sum of all absolute coefficients (Figure 4B). The thickness was the most important measure (60%), followed by the volumetric measures (29%), and finally functional connectivity (11%). The highest contributions came from four subtypes of thickness: bilateral patterns of cortical atrophy in temporal, sagittal and frontal areas (one subtype per hemisphere), and bilateral, opposite patterns of increased thickness (one subtype per hemisphere), Figure 4C. Two lateralized volumetric subtypes showed gray matter volume loss in the left motor, and right frontal areas as well as a gray volume increase in the left frontal and limbic regions. Finally, one functional subtype was very noisy and barely contributed to the model, while the other highlighted a connectivity subtype connecting the visual network with frontal areas.

Prediction of progression to dementia

We applied the HPS model to patients with MCI from the ADNI2 cohort, with the hypothesis that those would likely progress to AD dementia. The imaging sample for this experiment included the baseline structural and functional scans of all patients with MCI in the ADNI2 cohort ($N = 79$). We further stratified the patients with MCI into stable MCI (sMCI, $N = 37$), i.e. most recent clinical status is MCI with at least 36 months follow up, and progressors (pMCI, $N = 19$), i.e. individuals whose most recent known clinical status is AD dementia, with progression from MCI to AD dementia occurring within 37 months. The HPS model selected a subset of 10 subjects. Using the longitudinal follow-up clinical data provided by ADNI2, we found that 9 out of 10 of these subjects were pMCI (precision of 90%, specificity of 97%, sensitivity of 47%), compared to 34% pMCI in the whole MCI sample ($p < 0.001$), Figure 5A. Within the HPS subgroup, the time to progression from baseline to the first evaluation of AD dementia appeared uniformly distributed from 5 to 37 months, with 50% subjects progressing after 24 months (Figure 5C). In addition, 100% of the MCI participants flagged as HPS were positive for beta amyloid deposition with AV45 testing, compared to a 69% rate in the whole MCI sample ($p < 0.05$), Figure 5A. The rate of ApoE4 carriers in the HPS subsample was 78%, compared to 55% in the whole MCI group ($p > 0.05$), Figure 5A.

3. Discussion

The main goal of this work was to develop an imaging-based AD diagnosis with high precision and specificity. The proposed HPS approach did reach excellent performance in these respects, with 100% precision and specificity when distinguishing patients with AD dementia from CN participants and 90% precision, 98% specificity when predicting which MCI patients would progress to

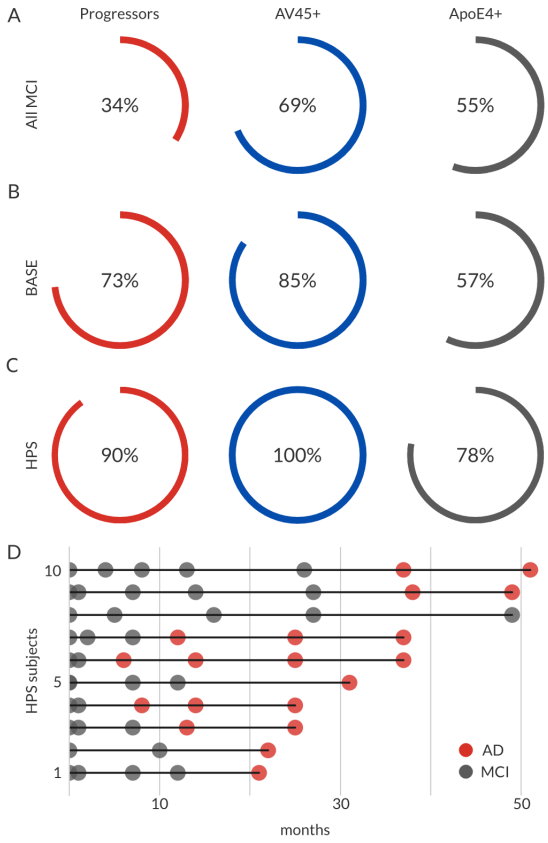

Figure 5: Statistic on the MCI showing the signature. Panel A shows the percentage of MCI who progress to AD, the percentage of subjects positive for beta amyloid deposits using the AV45 marker and the percentage of carriers of one or two copies of the ApoE4 allele for the entire MCI cohort. Panel B shows the same statistics for the selection of the base classifier while Panel C displays statistics for subjects flagged as HPS. Panel D shows the clinical status of each HPS subject over time from the baseline scan.

Table 1: Supervised classification of MCI progression to AD dementia using the ADNI database. Progression time was establish if the the subject progresses to AD status in the next 36 months. Significant improvement of our method compared to each paper for the adjusted precision (adjusted for a pMCI ratio of 34% comparable to our sample) and specificity are shown with $*$ for $p < 0.05$ and $**$ for $p < 0.001$) and conversely significant decrease in sensitivity of our method compared to each paper.

Author	N(sMCI, pMCI)	Precision	Precision (adjusted)	Specificity	Sensitivity
Dansereau et al. (This paper)	37, 19	**90%**	**90%**	**97%**	47%
Mathotaarachchi et al. (2017)	230, 43	51%	74%	87%*	71%*
Korolev et al. (2016)	120, 139	80%	65%*	76%**	83%*
Moradi et al. (2015)	100, 164	85%	63%*	74%**	87%**
Eskildsen et al. (2013)	134, 149	70%	52%**	68%**	66%
Wee et al. (2013)	111, 89	77%	68%*	84%**	64%
Gaser et al. (2013)	62, 133	90%	70%*	84%**	71%*
Davatzikos et al. (2011)	170, 69	57%	63%*	71%**	**95%****
Koikkalainen et al. (2011)	215, 154	66%	58%*	71%**	77%*
Misra et al. (2009)	76, 27	42%	51%**	60%**	80%*

dementia, up to three years before onset (see Table S2). These results represent a sizable and significant improvement in precision and specificity over the state of the art on this task, see Table 1. No data from MCI patients were used to train the model, which removes the possibility of a bias due to improper cross validation. The only HPS subject with MCI improperly classified as a progressor did have a series of 4 notes attached to his visits in the ADNI database, reporting on a decline in cognitive performance at each visit, and a marked decline at the last visit. This decline was still not severe enough for a diagnosis of AD dementia. The subject had no follow up available after 36 months, for unknown reason.

The high specificity of the HPS model came at the cost of a limited sensitivity: 38% when distinguishing patients with AD dementia from CN participants, and 47% when predicting which MCI patients would progress to dementia, which is significantly less than most recent published models, see Table 1. The HPS model is not designed to be sensitive, as it is trained to recognize a particular, homogenous brain signature present in only a fraction of the participants. The results of (Beach et al., 2012) suggest that only about half of patients diagnosed with AD dementia have clear AD brain markers post-mortem. The observed sensitivity of 38%- 47% is thus consistent with the idea that the HPS model is picking on a typical brain presentation of AD that is already present at the prodromal stage of the disease. Note that there was no need for patients with MCI to have as much absolute measures of atrophy as patients with AD dementia to

be recognized as HPS, as long as these patients presented with a similar spatial distribution of the atrophy, relative to other brain regions.

The anatomical features selected by the method were in line with recent subtyping works, e.g. (Hwang et al., 2015), showing predominant atrophy in the temporal lobe, as well as the temporo-parietal juncture, in particular. The functional maps were more difficult to interpret, and seemed more noisy. They still made a significant improvement in the performance of the HPS model. Because of the regularization in the logistic regression used to build the HPS model, features coming from different modalities did compete to be selected in the model. If redundant features existed, the ones with largest predictive power were selected by the classifier. This may explain why the selected functional subtypes did not involve the regions showing atrophy in the structural subtypes. We hypothesized that the HPS inferred from the AD vs CN prediction would also be useful to predict if a subject at the prodromal stage (MCI) would progress to dementia. Our results did validate this logic, but alternative strategies may be investigated in the future, e.g. training a model directly on the progressor vs stable MCI task.

A limitation of the present study was a moderate sample size, with $N = 56$ patients suffering from MCI. Although the ADNI is a large database, resting-state fMRI has only been added to the protocol in the latter stages of the study, ADNI GO and ADNI2. In addition, fMRI was only acquired on a third of participants, even after it was added to the protocol. Because of the early role of synaptic dysfunction in AD, and the potential ability of fMRI to capture such dysfunction, we wanted to build an anatomo-functional diagnostic tool. But this choice did limit the sample size of our study. In the future, we are planning to replicate our findings using structural measures only, so we can use the entirety of ADNI. Even with a larger sample size, another limitation of the ADNI dataset is that it does not reflect the diversity of cases observed in real-life clinical practice. Participants were in particular screened to exclude vascular dysfunction, which is a common comorbidity in AD. Resources with more inclusive enrollment criteria will become important in the future to better assess the generalizability of a biomarker-based AD diagnosis.

The most direct application of the HPS model is population enrichment for pharmaceutical clinical trials in AD (Woo et al., 2017; Mathotaarachchi et al., 2017). By recruiting almost exclusively patients who would normally progress to AD dementia, such enrichment would increase the effect size of the drug while reducing the sample size needed to demonstrate efficacy and therefore would also reduce the cost of the trial. The HPS brain signature is not shared among all the AD dementia population (making it a subtype), but is common enough to represent a substantial portion of participants of interest (about a third of AD dementia subjects and half of MCI progressors). An alternative enrichment strategy, more geared towards generalizability, would be to only exclude subjects that will very likely not progress to AD dementia. The HPS method thus brings us closer to precision medicine by proposing a middle ground between traditional clinical cohorts and an entirely individual medicine.

In this manuscript, we focused exclusively on two MRI modalities our ratio-

nale was that MRI is non-invasive and already widely used in patient care in elderly populations. Beta amyloid and tau PET imaging, by contrast, are more expensive and less available, while lumbar punctures are invasive. Nevertheless, it will be important in the future to see if a combination of PET imaging, blood tests looking for specific inflammatory proteins, cognitive scores, genetic factors, lifestyle factors, or others can help create multiple HPS that would in effect increase the sensitivity of the model for the early detection of Alzheimer's pathology.

4. Acknowledgments

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. 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 [1]. The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. This research was also supported by NIH grants P30 AG010129 and K01 AG030514.

The computational resources used to perform the data analysis were provided by Compute Canada[2] and CLUMEQ[3], which is funded in part by NSERC (MRS), FQRNT, and McGill University. This project was funded by NSERC grant number RN000028 and the Canadian Consortium on Neurodegeneration in Aging (CCNA), through a grant from the Canadian Institute of Health Research and funding from several partners including SANOFI-ADVENTIS R&D. CD is supported by a salary award from the Lemaire foundation. PB is supported by a salary award from "Fonds de recherche du Québec – Santé" and

[1] www.fnih.org
[2] https://computecanada.org/
[3] http://www.clumeq.mcgill.ca/

the Courtois Foundation.

5. Materials and methods

Dataset

All functional and structural data were obtained from the Alzheimer's Disease Neuroimaging Initiative 2 (ADNI2) sample, a longitudinal standardized acquisition including three populations: cognitively normal subjects, patients with mild cognitive impairment and patients with dementia due to AD. All participants gave their written informed consent to participate in the ADNI2 study, which was approved by the local ethics committee of participating institutions across North America. The consent form included data sharing with collaborators as well as secondary analysis. The present secondary analysis of the ADNI2 sample was approved by the local ethics committee at CRIUGM, University of Montreal, QC, Canada. All resting-state fMRI and structural scans were acquired on 3T Philips scanners with 8 channels. We performed analyses on the first usable scan (typically the baseline scan) from ADNI2.

The acquisition parameters were as follows: structural scan 170 slices, voxel size 1x1x1.2 mm3, matrix size 256x256, FOV 256 mm2, TR 6.8 s, TE 3.09 ms, FA 9 degrees. A functional scan of 7 min, 48 slices, voxel size 3.3x3.3x3.3 mm3, matrix size 64x64, FOV 212 mm2, TR 3 s, TE 30 ms, FA 80 degrees, No. volumes 140. For detailed information on the acquisition, see www.adni-info.org.

Extraction of functional features

Each fMRI dataset was corrected for slice timing; a rigid-body motion was then estimated for each time frame, both within and between runs, as well as between one fMRI run and the T1 scan for each subject (Collins et al., 1994). The T1 scan was itself non-linearly co-registered to the Montreal Neurological Institute (MNI) ICBM152 stereotaxic symmetric template (Fonov et al., 2011), using the CIVET pipeline (Ad-Dab'bagh et al., 2006a). The rigid-body, fMRI-to-T1 and T1-to-stereotaxic transformations were all combined to resample the fMRI in MNI space at a 3 mm isotropic resolution. To minimize artifacts due to excessive motion, all time frames showing a frame displacement, as defined in Power et al. (2012), greater than 0.5 mm were removed. An average residual frame displacement was also estimated after scrubbing for further group analyses. A minimum of 50 unscrubbed volumes per run was required for further analysis (13 subjects were rejected). The following nuisance covariates were regressed out from fMRI time series: slow time drifts (basis of discrete cosines with a 0.01 Hz highpass cut-off), average signals in conservative masks of the white matter and the lateral ventricles as well as the first principal components (accounting for 95% variance) of the six rigid-body motion parameters and their squares (Giove et al., 2009; Lund et al., 2006). The fMRI volumes were finally spatially smoothed with a 6 mm isotropic Gaussian blurring kernel. Datasets were preprocessed and analyzed using the NeuroImaging Analysis

Kit - NIAK - version 0.12.17 (http://niak.simexp-lab.org), under CentOS with Octave (http://gnu.octave.org) version 3.6.1 and the MINC toolkit (http://bic-mni.github.io/) version 0.3.18. Preprocessing of MRI data was executed in parallel on the Guillimin supercomputer (http://www.calculquebec.ca/en/resources/compute-servers/guillimin), using the pipeline system for Octave and Matlab - PSOM (Bellec et al., 2012). Seed-based fMRI connectivity maps were obtained using a functional brain template of 20 networks covering the entire brain. The Pearson's correlation between the average time series of each network and every voxel of the brain was computed to derive one functional connectivity map per network.

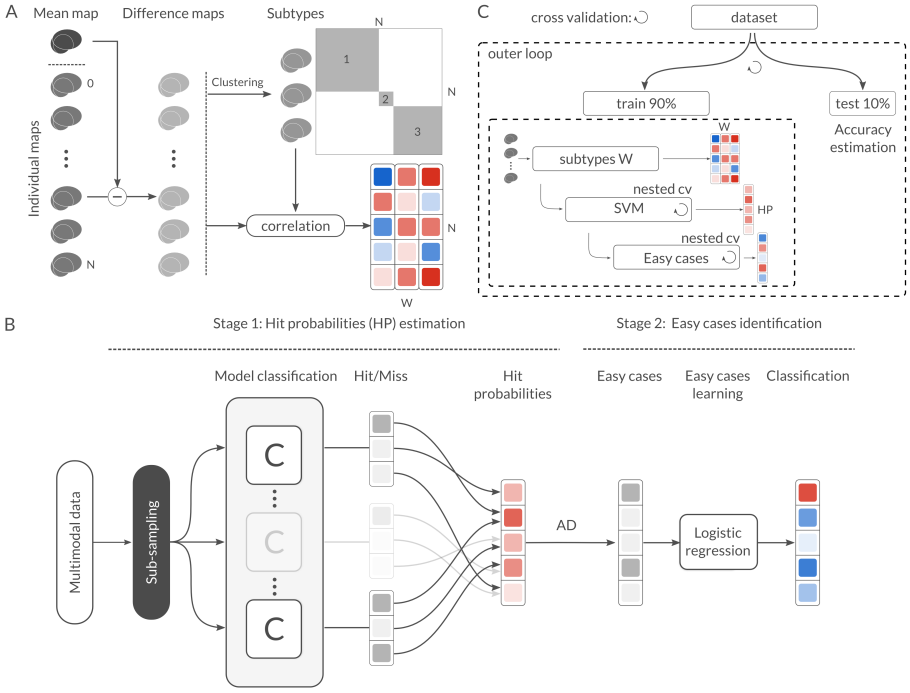

Figure 6: Panel A shows the feature extraction method called subtypes weights, Panel B framework workflow: stage 1 shows the hit probability computation based on random sub-sampling and stage 2 shows the training of dedicated classifier for each "high-confidence" signature. Panel C shows the nested cross-validation scheme used in this method.

Extraction of structural features

Native individual T1-weighted MRI scans were corrected for non-uniformity artifacts with the N3 algorithm (Sled et al., 1998). The corrected volumes were then masked for brain tissues (Smith, 2002) and registered into stereotaxic space (Collins et al., 1994). The registered, corrected images were segmented into gray matter (GM), white matter (WM), cerebrospinal fluid (CSF) and background using a neural net classifier (Tohka et al., 2004). The WM and GM surfaces

were extracted using the Constrained Laplacian-based Automated Segmentation with Proximities algorithm (Kim et al., 2005; MacDonald et al., 2000) and were resampled to a stereotaxic surface template to provide vertex based measures and lobar segmentation (Lyttelton et al., 2007). Cortical thickness was measured in native space using the linked distance between the two surfaces across 81,924 vertices (Im et al., 2008). Surface-based cortical thickness, as well as regional volume measures, were obtained using the structural MRI images processed using the CIVET 1.1.12 pipeline for each hemisphere as described in Ad-Dab'bagh et al. (2006b). The AAL template was applied on each hemisphere (40 regions per hemisphere) to extract the regional volumetric measures. The processing pipeline was executed on the Canadian Brain Imaging Network (CBRAIN) platform, a network of five imaging centers and eight High-Performance Computers for collaborative sharing and distributed processing of large MRI databases (Frisoni et al., 2011).

Multimodal imaging subtypes

We extracted subtypes that characterize the interindividual variability within the sample comprising CN and AD participants (at the time of scanning), independently for each type of measure (functional maps, cortical thickness maps, and volumetric maps). In order to reduce the impact of some factors of no interest that may influence the clustering procedure, we regressed out the age, sex, and average post-scrubbing frame displacement from individual maps, using a mass univariate linear regression model at each voxel. For each type of brain measure, we derived a spatial Pearson's correlation coefficient between all pairs of individual maps. This defined a subject x subject similarity matrix (of size 73 x 73), which was entered into a Ward hierarchical clustering procedure, as implemented in SciPy version 0.18.1 (Jones et al., 2001–; Walt et al., 2011). We arbitrarily selected three subgroups for each type of measure, based on a visual examination of the similarity matrix. For each type of measure, the average map of each subgroup defined a subtype. For each individual, we computed the spatial correlation of their map with each subtype. The resulting weight measures formed a matrix of size (number of subjects) x (number of subtypes), which was used as the feature space in the predictive models used throughout the rest of the methods. Note that this entire subtyping procedure, including regression of confounds, was latter entered in a cross-validation scheme to assess the performance of the predictive models.

Prediction of AD

The baseline prediction accuracy was obtained by training a SVM model with a linear kernel, as implemented in Scikit-learn Pedregosa et al. (2011) version 0.18. A tenfold cross-validation loop was used to estimate the performance of the trained model. Classes were balanced inversely, proportional to class frequencies in the input data for the training. A nested cross-validation loop was used (stratified shuffle split (50 splits, 20% test size)) for the grid search of the hyper-parameter C (grid was 10^{-2} to 10^{1} with 15 steps). Note that the C

parameter controlled how many misclassified examples the model will tolerate by adjusting the margin size. The model was evaluated using fMRI features only, sMRI features only, and the combination of fMRI and sMRI features.

Identifying easy cases

We randomly selected subsamples of the dataset (retaining 80% of participants in each subsample) to replicate the SVM training 100 times. For each 80% subsample, a separate SVM model was trained to predict the clinical labels (CN or AD), see Figure 6B. Note that the optimal C parameter was estimated once using the whole available sample, as described above, and used across all subsamples. This was done to avoid creating major uncontrolled algorithmic variations. The linear discriminating weights of the SVM were still optimized independently for each subsample. Predictions of clinical labels were then made on the remaining 20% of subjects, that were not used for training. For a given individual, the hit probability was calculated as the frequency of correct clinical classification across all available SVM replications where the test set included that individual. Easy cases were defined as individuals with 100

Predicting easy cases

We trained a logistic regression classifier Fan et al. (2008a) to predict the AD easy cases. The logistic regression was trained using a L1 regularization on the coefficients, see Figure 6B. Class weight was balanced inversely proportional to class frequencies in the input data. A stratified shuffle split (100 splits, 20% test size) was used to estimate the performance of the model for the grid search of the hyper-parameter C (grid was $10^{-0.2}$ to 10^{1} with 15 equal steps). In this case, the C parameter controlled the sparseness of the weights.

Cross-validation

A nested cross-validation was performed for accuracy estimation and parameters optimization. The outer loop used to estimate the generalizability of the framework was a ten-fold cross-validation scheme. Each training fold included the full multi-stage process of subtype extraction, SVM prediction of clinical labels, identification of HPS and prediction of HPS with logistic regression. Sensitivity (true positive rate, TP), specificity (true negative rate, TN) and precision ($TP/(TP + (1 - TN))$) of the diagnosis were estimated across all test folds, in the AD vs CN prediction. Cross-validation nested inside the outer loop was used to search for the optimal hyper-parameters, Figure 6C.

Highly predictive signature

The HPS was obtained by considering all subtypes associated with non-zero weights by the sparse logistic regression model in Figure 6C stage 2. All nonzero weights were considered as part of the signature and we used the corresponding map associated with each subtype weight.

Prediction of progression to dementia

The easy cases model was used to identify MCI patients who have a HPS of AD dementia. The imaging sample for this experiment included the baseline structural and functional scans of all patients with MCI in the ADNI2 cohort, with at least 36 months of follow-up ($N = 56$). We further stratified the patients with MCI into stable MCI (sMCI, $N = 37$)), i.e. latest clinical status is MCI, and progressors (pMCI, $N = 19$), i.e. individuals whose most recent known clinical status is AD dementia, with progression from MCI to AD dementia occurring within 36 months. Note that no AV45 imaging data or genetic data, nor any data from the MCI cohort, were used to build the HPS model.

Statistical test of differences in model performance

We generated a confidence interval on the performance (i.e. precision, specificity and sensitivity) of a given model using a Monte-Carlo simulation. Taking the observed sensitivity and specificity, and using similar sample size to our experiment, we replicated the number of true and false positive detection 100000 times using independent Bernoulli variables, and derived replications of precision, specificity and sensitivity. By comparing these replications to the sensitivity, specificity and precision observed in other models, we estimated a p-value for differences in model performance (Phipson and Smyth, 2010). A p-value smaller than 0.05 was interpreted as evidence of a significant difference in performance, and 0.001 as a strong evidence. This approach was first used in Figure 3 to contrast the performance of the HPS model to the baseline (SVM) model, both for AD vs CN and MCI progressor vs stable, as well as contrasting the performance of multimodal (fMRI+sMRI) model vs models using only fMRI or sMRI features. The same approach was used to contrast our proposed model for MCI progressor vs stable with results from the literature, in Table 1. Note that, reflecting our hypotheses regarding the behaviour of the HPS model, the tests were one-sided for increase in specificity and precision, and one-sided for decrease in sensitivity.

Statistical test of enrichment

The HPS model was used to select a subset of the MCI population. We tested statistically if this subgroup was enriched for (1) progression to dementia; (2) AV45+, and; (3) ApoE4+. We implemented for this purpose a Monte-Carlo simulation, where we selected 100000 random subgroups out of the original MCI sample. By comparing the proportion of progressors (respectively AV45+ and ApoE4+) in these null replications to the actual observed values in the HPS subgroup, we estimated a p-value (Phipson and Smyth, 2010) (one sided for increase). A p-value smaller than 0.05 was interpreted as evidence of a significant enrichment, and 0.001 as a strong evidence.

Public code and data

The code used in this experiment is available on a GitHub repository at the following URL[4]. An IPython Notebook is also provided with all of the figure generation scripts. Scikit-learn Pedregosa et al. (2011) version 0.18 was used for most of the machine learning algorithms and Nilearn Abraham et al. (2014) version 0.2.6 for visualization purposes.

ADNI dataset

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).

References

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., Varoquaux, G., 2014. Machine learning for neuroimaging with scikit-learn. Frontiers in Neuroinformatics 8, 14.
URL `http://journal.frontiersin.org/article/10.3389/fninf.2014.00014`

Ad-Dab'bagh, Y., Einarson, D., Lyttelton, O., Muehlboeck, J. S., Mok, K., Ivanov, O., Vincent, R. D., Lepage, C., Lerch, J., Fombonne, E., Evans, A. C., 2006a. The CIVET Image-Processing environment: A fully automated comprehensive pipeline for anatomical neuroimaging research. In: Corbetta, M. (Ed.), Proceedings of the 12th Annual Meeting of the Human Brain Mapping Organization. Neuroimage, Florence, Italy.

Ad-Dab'bagh, Y., Einarson, D., Lyttelton, O., Muehlboeck, J. S., Mok, K., Ivanov, O., Vincent, R. D., Lepage, C., Lerch, J., Fombonne, E., Evans, A. C., 2006b. The CIVET Image-Processing Environment: A Fully Automated Comprehensive Pipeline for Anatomical Neuroimaging Research. In: Corbetta, M. (Ed.), Proceedings of the 12th Annual Meeting of the Human Brain Mapping Organization. Neuroimage, Florence, Italy.

Arbabshirani, M. R., Plis, S., Sui, J., Calhoun, V. D., 2017. Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. NeuroImage 145, 137–165.

[4] `https://github.com/simexp/hpc`

Au, R., Piers, R. J., Lancashire, L., 2015. Back to the future: Alzheimer's disease heterogeneity revisited. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring 1 (3), 368.

Beach, T. G., Monsell, S. E., Phillips, L. E., Kukull, W., 2012. Accuracy of the clinical diagnosis of alzheimer disease at national institute on aging alzheimer disease centers, 2005–2010. Journal of neuropathology & experimental neurology 71 (4), 266–273.

Bellec, P., Benhajali, Y., Carbonell, F., Dansereau, C., Albouy, G., Pelland, M., Craddock, C., Collignon, O., Doyon, J., Stip, E., et al., 2015. Impact of the resolution of brain parcels on connectome-wide association studies in fmri. NeuroImage 123, 212–228.

Bellec, P., Lavoie-Courchesne, S., Dickinson, P., Lerch, J. P., Zijdenbos, A. P., Evans, A. C., 2012. The pipeline system for Octave and Matlab (PSOM): a lightweight scripting framework and execution engine for scientific workflows. Frontiers in neuroinformatics 6.
URL http://dx.doi.org/10.3389/fninf.2012.00007

Cheng, B., Liu, M., Suk, H.-I., Shen, D., Zhang, D., Initiative, A. D. N., et al., 2015a. Multimodal manifold-regularized transfer learning for mci conversion prediction. Brain imaging and behavior 9 (4), 913–926.

Cheng, B., Liu, M., Zhang, D., Munsell, B. C., Shen, D., 2015b. Domain transfer learning for mci conversion prediction. IEEE Transactions on Biomedical Engineering 62 (7), 1805–1817.

Collins, D. L., Neelin, P., Peters, T. M., Evans, A. C., 1994. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. Journal of computer assisted tomography 18 (2), 192–205.
URL http://view.ncbi.nlm.nih.gov/pubmed/8126267

Cortes, C., Vapnik, V., Sep. 1995. Support-vector networks. Machine Learning 20 (3), 273–297.
URL http://dx.doi.org/10.1007/BF00994018

Dubois, B., Feldman, H. H., Jacova, C., Dekosky, S. T., Barberger-Gateau, P., Cummings, J., Delacourte, A., Galasko, D., Gauthier, S., Jicha, G., Meguro, K., O'brien, J., Pasquier, F., Robert, P., Rossor, M., Salloway, S., Stern, Y., Visser, P. J., Scheltens, P., Aug. 2007. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. Lancet neurology 6 (8), 734–746.
URL http://dx.doi.org/10.1016/S1474-4422(07)70178-3

Eskildsen, S. F., Coupé, P., Garc'ıa-Lorenzo, D., Fonov, V., Pruessner, J. C., Collins, D. L., Jan. 2013. Prediction of Alzheimer's disease in subjects with mild cognitive impairment from the ADNI cohort using patterns of cortical thinning. NeuroImage 65, 511–521.
URL http://dx.doi.org/10.1016/j.neuroimage.2012.09.058

Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., Lin, C.-J., 2008a. Liblinear: A library for large linear classification. The Journal of Machine Learning Research 9, 1871–1874.

Fan, Y., Batmanghelich, N., Clark, C. M., Davatzikos, C., Initiative, A. D. N., et al., 2008b. Spatial patterns of brain atrophy in mci patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. Neuroimage 39 (4), 1731–1743.

Fodero-Tavoletti, M. T., Okamura, N., Furumoto, S., Mulligan, R. S., Connor, A. R., McLean, C. A., Cao, D., Rigopoulos, A., Cartwright, G. A., OKeefe, G., et al., 2011. 18f-thk523: a novel in vivo tau imaging ligand for alzheimers disease. Brain 134 (4), 1089–1100.

Fonov, V., Evans, A. C., Botteron, K., Almli, C. R., McKinstry, R. C., Collins, D. L., Brain Development Cooperative Group, Jan. 2011. Unbiased average age-appropriate atlases for pediatric studies. NeuroImage 54 (1), 313–327.
URL http://dx.doi.org/10.1016/j.neuroimage.2010.07.033

Frisoni, G. B., Redolfi, A., Manset, D., Rousseau, M.-E., Toga, A., Evans, A. C., Jul. 2011. Virtual imaging laboratories for marker discovery in neurodegenerative diseases. Nature Reviews Neurology 7 (8), 429–438.
URL http://dx.doi.org/10.1038/nrneurol.2011.99

Giove, F., Gili, T., Iacovella, V., Macaluso, E., Maraviglia, B., Oct. 2009. Images-based suppression of unwanted global signals in resting-state functional connectivity studies. Magnetic resonance imaging 27 (8), 1058–1064.
URL http://dx.doi.org/10.1016/j.mri.2009.06.004

Hwang, J., Kim, C. M., Jeon, S., Lee, J. M., Hong, Y. J., Roh, J. H., Lee, J.-H., 2015. Prediction of alzheimers disease pathophysiology based on cortical thickness patterns. Alzheimer's & Dementia: The Journal of the Alzheimer's Association 11 (7), P541.

Im, K., Lee, J.-M., Lyttelton, O., Kim, S. H., Evans, A. C., Kim, S. I., 2008. Brain size and cortical structure in the adult human brain. Cerebral Cortex 18 (9), 2181–2191.

Jellinger, K. A., et al., 2014. Neuropathology of dementia disorders. J. Alz. Dis. Parkinsonism 4, 135.

Jones, E., Oliphant, T., Peterson, P., et al., 2001–. SciPy: Open source scientific tools for Python. [Online; accessed ¡today¿].
URL http://www.scipy.org/

Karas, G., Scheltens, P., Rombouts, S., Visser, P., Van Schijndel, R., Fox, N., Barkhof, F., 2004. Global and local gray matter loss in mild cognitive impairment and alzheimer's disease. Neuroimage 23 (2), 708–716.

Kim, J. S., Singh, V., Lee, J. K., Lerch, J., Ad-Dab'bagh, Y., MacDonald, D., Lee, J. M., Kim, S. I., Evans, A. C., 2005. Automated 3-d extraction and evaluation of the inner and outer cortical surfaces using a laplacian map and partial volume effect classification. Neuroimage 27 (1), 210–221.

Korolev, I. O., Symonds, L. L., Bozoki, A. C., Initiative, A. D. N., et al., 2016. Predicting progression from mild cognitive impairment to alzheimer's dementia using clinical, mri, and plasma biomarkers via probabilistic pattern classification. PloS one 11 (2), e0138866.

Lerch, J. P., Pruessner, J. C., Zijdenbos, A., Hampel, H., Teipel, S. J., Evans, A. C., Jul. 2005. Focal Decline of Cortical Thickness in Alzheimer's Disease Identified by Computational Neuroanatomy. Cerebral Cortex 15 (7), 995–1001.
URL http://dx.doi.org/10.1093/cercor/bhh200

Lund, T. E., Madsen, K. H., Sidaros, K., Luo, W.-L., Nichols, T. E., Jan. 2006. Non-white noise in fMRI: does modelling have an impact? NeuroImage 29 (1), 54–66.
URL http://dx.doi.org/10.1016/j.neuroimage.2005.07.005

Lyttelton, O., Boucher, M., Robbins, S., Evans, A., 2007. An unbiased iterative group registration template for cortical surface analysis. Neuroimage 34 (4), 1535–1544.

MacDonald, D., Kabani, N., Avis, D., Evans, A. C., 2000. Automated 3-d extraction of inner and outer surfaces of cerebral cortex from mri. NeuroImage 12 (3), 340–356.

Marklund, N., Blennow, K., Zetterberg, H., Ronne-Engström, E., Enblad, P., Hillered, L., Jun. 2009. Monitoring of brain interstitial total tau and beta amyloid proteins by microdialysis in patients with traumatic brain injury. J. Neurosurg. 110 (6), 1227–1237.

Mathotaarachchi, S., Pascoal, T. A., Shin, M., Benedet, A. L., Kang, M. S., Beaudry, T., Fonov, V. S., Gauthier, S., Rosa-Neto, P., Initiative, A. D. N., et al., 2017. Identifying incipient dementia individuals using machine learning and amyloid imaging. Neurobiology of Aging.

McConathy, J., Sheline, Y. I., 2015. Imaging biomarkers associated with cognitive decline: a review. Biological psychiatry 77 (8), 685–692.

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Kawas, C. H., Klunk, W. E., Koroshetz, W. J., Manly, J. J., Mayeux, R., et al., 2011. The diagnosis of dementia due to alzheimers disease: Recommendations from the national institute on aging-alzheimers association workgroups on diagnostic guidelines for alzheimer's disease. Alzheimer's & dementia 7 (3), 263–269.

Moradi, E., Pepe, A., Gaser, C., Huttunen, H., Tohka, J., Initiative, A. D. N., et al., 2015. Machine learning framework for early mri-based alzheimer's conversion prediction in mci subjects. Neuroimage 104, 398–412.

Orban, P., Tam, A., Urchs, S., Savard, M., Madjar, C., Badhwar, A., Dansereau, C., Vogel, J., Shmuel, A., Dagher, A., Villeneuve, S., Poirier, J., Rosa-Neto, P., Breitner, J., Bellec, P., , , 2017. Subtypes of functional brain connectivity as early markers of neurodegeneration in alzheimer's disease. bioRxiv.
URL https://www.biorxiv.org/content/early/2017/09/28/195164

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 12, 2825–2830.

Phipson, B., Smyth, G. K., 2010. Permutation p-values should never be zero: calculating exact p-values when permutations are randomly drawn. Statistical applications in genetics and molecular biology 9 (1).

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., Petersen, S. E., Feb. 2012. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. NeuroImage 59 (3), 2142–2154.
URL http://dx.doi.org/10.1016/j.neuroimage.2011.10.018

Querbes, O., Aubry, F., Pariente, J., Lotterie, J.-A., Démonet, J.-F., Duret, V., Puel, M., Berry, I., Fort, J.-C., Celsis, P., et al., 2009. Early diagnosis of alzheimer's disease using cortical thickness: impact of cognitive reserve. Brain 132 (8), 2036–2047.

Rabinovici, G. D., Carrillo, M. C., Forman, M., DeSanti, S., Miller, D. S., Kozauer, N., Petersen, R. C., Randolph, C., Knopman, D. S., Smith, E. E., et al., 2017. Multiple comorbid neuropathologies in the setting of alzheimer's disease neuropathology and implications for drug development. Alzheimer's & Dementia: Translational Research & Clinical Interventions 3 (1), 83–91.

Rathore, S., Habes, M., Iftikhar, M. A., Shacklett, A., Davatzikos, C., 2017. A review on neuroimaging-based classification studies and associated feature extraction methods for alzheimer's disease and its prodromal stages. NeuroImage.

Sled, J. G., Zijdenbos, A. P., Evans, A. C., 1998. A nonparametric method for automatic correction of intensity nonuniformity in mri data. IEEE transactions on medical imaging 17 (1), 87–97.

Smith, S. M., Nov. 2002. Fast robust automated brain extraction. Hum. Brain Mapp. 17 (3), 143–155.
URL http://dx.doi.org/10.1002/hbm.10062

Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., Iwatsubo, T., Jack, C. R., Kaye, J., Montine, T. J., Park, D. C., Reiman, E. M., Rowe, C. C., Siemers, E., Stern, Y., Yaffe, K., Carrillo, M. C., Thies, B., Morrison-Bogorad, M., Wagster, M. V., Phelps, C. H., May 2011. Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's & Dementia 7 (3), 280–292.
URL http://dx.doi.org/10.1016/j.jalz.2011.03.003

Tohka, J., Zijdenbos, A., Evans, A., 2004. Fast and robust parameter estimation for statistical partial volume models in brain mri. Neuroimage 23 (1), 84–97.

Toussaint, P.-J., Perlbarg, V., Bellec, P., Desarnaud, S., Lacomblez, L., Doyon, J., Habert, M. O., Benali, H., 2012. Resting state fdg-pet functional connectivity as an early biomarker of alzheimer's disease using conjoint univariate and independent component analyses. NeuroImage 63 2, 936–46.

Walt, S. v. d., Colbert, S. C., Varoquaux, G., 2011. The numpy array: a structure for efficient numerical computation. Computing in Science & Engineering 13 (2), 22–30.

Woo, C.-W., Chang, L. J., Lindquist, M. A., Wager, T. D., 2017. Building better biomarkers: brain models in translational neuroimaging. Nature neuroscience 20 (3), 365–377.

Xu, L., Wu, X., Chen, K., Yao, L., 2015. Multi-modality sparse representation-based classification for alzheimer's disease and mild cognitive impairment. Computer methods and programs in biomedicine 122 (2), 182–190.

Zheng, W., Yao, Z., Hu, B., Gao, X., Cai, H., Moore, P., 2015. Novel cortical thickness pattern for accurate detection of alzheimers disease. Journal of Alzheimer's Disease 48 (4), 995–1008.

Zhu, X., Suk, H.-I., Shen, D., 2014. A novel matrix-similarity based loss function for joint regression and classification in ad diagnosis. NeuroImage 100, 91–105.

Zu, C., Jie, B., Liu, M., Chen, S., Shen, D., Zhang, D., Initiative, A. D. N., et al., 2016. Label-aligned multi-task feature learning for multimodal classification of alzheimers disease and mild cognitive impairment. Brain imaging and behavior 10 (4), 1148–1159.

Supplementary Material – A brain signature highly predictive of future progression to Alzheimers dementia

Table S2: Performance of the models. Prec: precision, Spec: specificity, Sens: sensitivity and N: number of selected subjects.

Modality	Algo	Contrast	Prec (%)	Spec (%)	Sens (%)	N
fMRI	Base	CN/AD	38.10	46.94	66.67	42
	HPS		60	95.92	12.5	5
sMRI	Base		66.67	83.67	66.67	24
	HPS		87.50	97.96	29.17	8
fMRI+sMRI	Base		69.57	85.71	66.67	23
	HPS		100	100	37.50	9
fMRI+sMRI	Base	sMCI/pMCI	73.33	89.19	57.89	15
	HPS		90	97.3	47.37	10

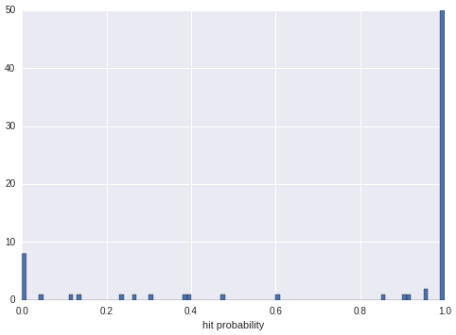

Figure S1: Hit-probability distribution obtained from replicating the SVM training 100 times from 80% of the training set.