Transmodal Learning of Functional Networks for Alzheimer's Disease Prediction

Mehdi Rahim, Bertrand Thirion, Claude Comtat, Gaël Varoquaux

Transmodal Learning of Functional Networks for Alzheimer's Disease Prediction

Mehdi Rahim [*], Bertrand Thirion [†], Claude Comtat [‡], Gaël Varoquaux, for the Alzheimer's Disease Neuroimaging Initiative [§]

Abstract

Functional connectivity describes neural activity from resting-state functional magnetic resonance imaging (rs-fMRI). This noninvasive modality is a promising imaging biomarker of neurodegenerative diseases, such as Alzheimer's disease (AD), where the connectome can be an indicator to assess and to understand the pathology. However, it only provides noisy measurements of brain activity. As a consequence, it has shown fairly limited discrimination power on clinical groups. So far, the reference functional marker of AD is the fluorodeoxyglucose positron emission tomography (FDG-PET). It gives a reliable quantification of metabolic activity, but it is costly and invasive. Here, our goal is to analyze AD populations solely based on rs-fMRI, as functional connectivity is correlated to metabolism. We introduce *transmodal learning*: leveraging a prior from one modality to improve results of another modality on different subjects. A metabolic prior is learned from an independent FDG-PET dataset to improve functional

[*]- M. Rahim is with Parietal project team - INRIA Saclay and with IMIV team - CEA Saclay DRF/I2BM/NeuroSpin and SHFJ. Paris-Saclay University. France. (e-mail: mehdi.rahim@cea.fr)

[†]- B. Thirion and G. Varoquaux are with Parietal project team - INRIA Saclay and CEA Saclay DRF/I2BM/NeuroSpin. Paris-Saclay University. France.

[‡]- C. Comtat is with IMIV team - CEA Saclay DRF/I2BM/SHFJ. Paris-Saclay University. France.

[§]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

connectivity-based prediction of AD. The prior acts as a regularization of connectivity learning and improves the estimation of discriminative patterns from distinct rs-fMRI datasets. Our approach is a two-stage classification strategy that combines several seed-based connectivity maps to cover a large number of functional networks that identify AD physiopathology. Experimental results show that our transmodal approach increases classification accuracy compared to pure rs-fMRI approaches, without resorting to additional invasive acquisitions. The method successfully recovers brain regions known to be impacted by the disease.

1 Introduction

Brain imaging can probe the signatures, anatomical or functional, of brain diseases. For Alzheimer's Disease (AD), anatomical measurements on Magnetic Resonance Images (MRIs) such as the hippocampus volume [1] or the cortical thickness [2, 3] are accurate biomarkers that help to distinguish AD subjects from Mild Cognitive Impairment (MCI) subjects who convert later to AD [4].

Another imaging modality, Positron Emission Tomography (PET), measures brain biochemical and functional dynamics according to specific radiotracers. For example, the Pittsburgh compound B (PiB) radiotracer quantifies beta-amyloid plaques deposition that is at the root of AD [5]. FDG-PET (fluorodeoxyglucose PET) measures quantitatively brain metabolic activity and is considered as sufficiently accurate biomarker of AD. FDG-PET analysis has shown that a specific pattern of decrease of metabolic activity characterizes AD subjects with an accuracy of 90% [6, 7] However, PET imaging is costly and implies exposing the subject to radiations.

More recently, studies have shown promising use of resting-state functional MRI (rs-fMRI) as a biomarker of AD [8], [9]. Rs-fMRI is an easy-to-implement imaging protocol. It reveals functional interactions between brain networks that predict brain states [10], via the intrinsic functional connectivity (FC) estimated from the blood-oxygen-level dependent (BOLD) signal. As studied in [11], [12] and [13], AD is characterized by widespread decreases in connectivity, especially in the default mode network (DMN). Although functional connectivity brings a supplementary information on AD by mapping functional regions that share some common dynamics, its sensitivity to classify correctly AD subjects [14] is lower than the anatomical MRI features and FDG-PET. fMRI has a low signal-to-noise ratio (SNR), measures of functional connectivity are often limited by the quantity and quality of data. Reproducibility of the connectome is another major limitation. Indeed,

many studies ([15], [16]) have shown that good reproducibility is achievable only with longer time-series and after several re-test sessions.

Recent studies correlating FDG-PET to rs-fMRI have shown that functional connectivity at a regional level reflects the metabolic activity. In [17], the authors have presented a simultaneous PET/MR study on subjects watching visual stimuli. Correlations were found between metabolism and seed-based functional connectivity in visual areas during stimulation. Other works [18, 19] have shown significant correlation between the metabolism and the functional connectivity at the DMN level, in particular in the precuneus and the posterior cingulate cortex. Our contribution builds on this relationship between metabolism and connectivity: we conjecture that patterns specific to AD in fMRI and PET are equivalent [20]. We propose a *transmodal learning* framework that estimates a predictive model classifying AD from rs-fMRI –noninvasive but weakly accurate– informed by an FDG-PET (invasive but accurate) discriminative pattern. Unlike multi-modal learning that combines different modalities observed in each subject, transmodal learning relies on a metabolic model to inform a connectivity model. This transfer implies that our connectivity-based model can be used on rs-fMRI data acquired independently from the FDG-PET dataset used to build the metabolic prior. As there have already been many studies of AD with FDG-PET, a lot of information can be leveraged to improve connectivity-based prediction. Our goal is to have a more accurate discriminant model on non-invasive rs-fMRI to avoid as much as possible the use of an invasive modality. The proposed model addresses the issue of the arbitrary selection of a reference region of interest (ROI) when computing functional connectivity. Seed-based correlations are computed according to ROIs extracted from a functional atlas, followed by a model that stacks their predictions. Such an atlas may include ROIs that do not exhibit high correlation between connectivity and metabolism. However, the stacking model that combines them will select the most relevant ROIs. Resulting predictor, based on multiple ROIs, gives better accuracy than a single ROI.

Prior work in the neuroimaging literature has relied on informing a given imaging modality with a prior extracted from a different but related modality. In particular, potential correlations between functional and anatomical characteristics have been considered. In [21], functional connectivity matrix was learned from rs-fMRI with a structural-connectivity constraint from diffusion weighted MRI (dMRI). [22] proposed a unsupervised decomposition of resting-state time-series from rs-fMRI. This decomposition is constrained by a fiber connection pattern extracted from dMRI. While [23] introduced a joint functional/structural connectivity estimation that helped to identify joint connectivity variations in schizophrenia. However, these two modalities

should be combined with caution. Indeed, [24] highlighted possible inconsistencies between fMRI and dMRI. From a methodological standpoint, our work is related to transfer learning methods [25] that apply a model learned from a dataset (or modality) A to a dataset B. In AD classification, [26] has used a transfer-learning scheme called *auxilary model* to enhance mild-cognitive-converters prediction from FDG-PET, anatomical MRI, and protein biomarkers. The main novelty of our approach is that we use the learned model as a constraint to the classification training.

A preliminary work has been presented in [27], where we have shown the feasibility of the approach on a smaller dataset. Here, after detailing the transmodal learning model, we extensively validate and analyze it: *i)* we compare this approach to classical approaches; *ii)* we assess the impact of parameter choice; *iii)* we extend the approach to a multi-modal dataset, where we show that the metabolism-informed connectivity yields more accurate classification of AD.

In section 2, we detail the transmodal learning framework for the connectivity-based prediction. Then we present in section 3 data used in our experiments that have been extracted from 211 subjects across 694 fMRI scans of the ADNI database. The results shown in section 4 demonstrate the efficiency of the approach compared to classical connectivity-based classification.

2 Methods

Here we introduce an enhanced predictive model of AD from functional connectivity computed from resting-state fMRI. This model exploits accurate and discriminative knowledge from previous FDG-PET studies by assuming that changes in metabolic activity yield information comparable to functional connectivity alterations.

We propose a voxel-based classifier that handles the connectivity-metabolism relationship. For this purpose, region-to-voxel connectivities are computed according to several regions of interest (ROIs). This leads to one predictive model per ROI, hence these models have to be combined at a later stage. This problem is often depicted as a multi-source learning and several solutions have been proposed such as multiple kernel learning (MKL) [28]. In the MKL setting, a kernel is computed for each source. Then a linear combination of the kernels is learned to optimize the classification of the sources. Another solution is to train a classifier on each source, and then to use majority voting to determine the final class assignment. We use a stacking approach, relying on a second-stage classification model that takes as input model predictions from each ROI. This approach makes no assumptions on

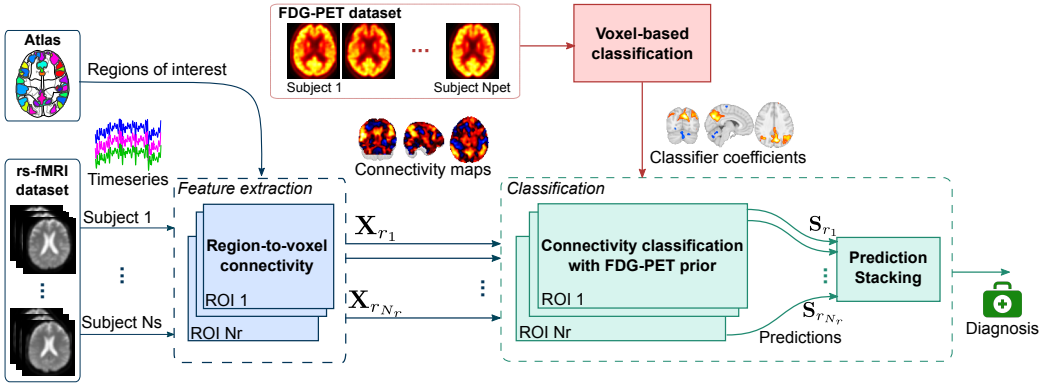


Figure 1: **Overview of the proposed connectivity classification framework.** First, region-to-voxel connectivity is computed from rs-fMRI timeseries according to each ROI of a brain atlas. This yields N_r connectivity maps for each subject. The FDG-PET prior is a vector of weights learned from a classification model on a distinct and larger dataset than the rs-fMRI dataset. These weights are integrated to regularize a linear classifier on the ROIs connectivity maps. Finally, predictions of all ROIs are concatenated in a matrix which is given as input to a stacking model to predict the clinical group.

the classifiers that should be used, and is well suited to heterogeneous data. It can also be used to recover the discriminant features and spatial maps, as it operates at the voxel-level with a linear classifier.

2.1 Transmodal Learning Framework Overview

The proposed pipeline is depicted in Fig. 1. The inputs are 4D acquisitions of resting-state fMRI, and a set of regions of interest (ROIs) from which the connectivity is estimated. First, the prior is estimated from 3D FDG-PET images of metabolism at the voxel-level with a linear classifier. The resting-state fMRI features are obtained by computing region-to-voxel connectivity maps from fMRI data. Region-to-voxel connectivity maps depend strongly on the the specific choice of seed region [29]. To avoid relying on a single ROI, we propose to combine a set of ROIs extracted from a brain atlas. Our goal is to aggregate discriminative functional-connectivity features from an atlas covering various functional networks. We will discuss in section 4.4 which atlas should be selected. Then, an enhanced regression model informed by the FDG-PET prior is estimated; this yields one model per ROI. Finally, the global fMRI-based classification is performed via another classifier that estimates the target from the stacked predictions of each ROI-based model.

Table 1: Notations used in the method description.

N_v	number of voxels of the entire brain.
N_r	number of ROIs.
N_s	number of subjects of the rs-fMRI dataset.
$\mathcal{A}$	brain atlas: set of N_r brain regions r_i, such that: $\mathcal{A} = \{ r_i \}$ with $1 \leq i \leq N_r$.
$\mathbf{ts}_{v_j}^{(s_k)}$	time-series at a voxel v_j in subject s_k.
$\mathbf{ts}_{r_i}^{(s_k)}$	averaged time-series across all voxels of ROI r_i in subject s_k.
$fc_{r_i,v_j}^{(s_k)}$	connectivity at voxel v_j according to ROI r_i in subject s_k.
$fc_{r_{i_1},r_{i_2}}^{(s_k)}$	connectivity between ROI r_{i_1} and ROI r_{i_2} in subject s_k.
$\mathbf{fc}_{r_i}^{(s_k)}$	connectivity map (N_v voxels) in subject s_k according to ROI r_i.
$\mathbf{X}_{r_i}$	connectivities of all subjects according to ROI r_i, its dimension is (N_s, N_v).
$\mathbf{y}$	subject binary class (diagnosis) vector, its dimension is N_s.
$N_{s_{\text{pet}}}$	number of subjects of the FDG-PET dataset.
$\mathbf{X}_{\text{pet}}$	feature matrix of the FDG-PET dataset, its dimension is $(N_{s_{\text{pet}}}, N_v)$.
$\mathbf{y}_{\text{pet}}$	diagnosis vector of dimension $N_{s_{\text{pet}}}$.
$\mathbf{w}_{\text{prior}}$	linear classifier coefficients of the FDG-PET dataset, its dimension is N_v.
$\mathbf{w}_{r_i}$	linear classifier coefficients from the connectivity values associated with the seed region r_i.

Table 1 gives the notations that we use for the mathematical formulations.

2.2 Functional Connectivity Features

There are generally two ways to compute FC:

Region-to-voxel Region-to-voxel connectivity measures correlations from a given seed ROI to the whole brain. It gives for each ROI a connectivity map at the voxel level. Choosing a seed ROI is difficult, as it changes vastly this connectivity map. We overcome this limitation by considering a set of ROIs. Typically, each region of a given brain atlas $\mathcal{A}$ is used as a seed ROI for a connectivity map. For a subject s_k, we define the region-to-voxel connectivity $fc^{(s_k)}$ between any voxel and the region of interest r_i as the Pearson correlation between their time-series [30]. The correlation values are then normalized using Fisher's Z-transformation. It converts correlations to normally distributed values for a better comparison.

Region-to-region Region-to-region connectivity $fc^{(s_k)}_{r_{i_1},r_{i_2}}$ between ROI r_{i_1} and ROI r_{i_2} is the correlation between their respective averaged timeseries. The correlation is also normalized with Fisher's Z-transformation. This measurement on several ROIs (or atlas) yields a connectivity matrix (connectome) characterizing a subject [31]. It is not used in our approach, which relies on only region-to-voxel connectivities, but we will compare the two sets of features in the experiments.

The connectivity map $\mathbf{fc}^{(s_k)}$ associated with a ROI r_i is a vector of dimension N_v that represents the correlations of all brain voxels

$$\mathbf{fc}^{(s_k)}_{r_i} = \left(fc^{(s_k)}_{r_i,v_1},\ \dots, fc^{(s_k)}_{r_i,v_j},\ \dots, fc^{(s_k)}_{r_i,v_{N_v}} \right). \tag{1}$$

We define the feature matrix $\mathbf{X}_{r_i}$ of dimension (N_s, N_v) associated with ROI r_i as:

$$\mathbf{X}_{r_i} = \left(\mathbf{fc}^{(s_1)}_{r_i},\ \dots, \mathbf{fc}^{(s_k)}_{r_i},\ \dots, \mathbf{fc}^{(s_{N_s})}_{r_i} \right). \tag{2}$$

This matrix will be used for the diagnosis prediction. We compute one classification model per ROI. Then the diagnosis is learned from the models predictions according to all ROIs.

2.3 Data-Driven Metabolic Prior Integration

Rather than a multi-modal PET-fMRI prediction in each subject, we derive independently a population-level PET prior, thus avoiding the additional requirement of one PET acquisition per subject for diagnosis. The estimation of a connectivity-based classification model is regularized by a learned metabolic prior, which involves a complementary coupling parameter to adapt this prior.

2.3.1 Prior Estimation

The metabolic prior is composed of discriminative coefficients of a linear model learned from $N_{s_{\text{pet}}}$ FDG-PET images of a dataset distinct from the rs-fMRI dataset. It is the core of the transmodal approach that we propose. We define a feature matrix $\mathbf{X}_{\text{pet}}$ representing the metabolism in each voxel and a label vector $\mathbf{y}_{\text{pet}}$ representing the target (diagnosis). A linear model is calculated on this matrix $\mathbf{X}_{\text{pet}}$, where the model coefficients $\hat{\mathbf{w}}_{\text{prior}}$ are estimated as follows:

$$\hat{\mathbf{w}}_{\text{prior}} = \underset{\mathbf{w}\in\mathbb{R}^{N_v}}{\operatorname{argmin}} \ \mathcal{L}(\mathbf{X}_{\text{pet}}\mathbf{w}, \mathbf{y}_{\text{pet}}) + \Omega(\mathbf{w}), \tag{3}$$

where $\mathcal{L}$ is the loss function of the prediction which can be seen as a data-fidelity term. Ω is a regularization term. We keep this general formulation of a linear classifier since we are free to choose which classifier to use. In out experiments, we use a ℓ_2 regularized logistic regression classifier to generate the metabolic prior $\mathbf{w}_{\text{prior}} \in \mathbb{R}^{N_v}$. It will be integrated into the functional connectivity classification.

2.3.2 Transmodal Classification

The integration of the metabolic prior into functional connectivity-based classification is done at the voxel level, by assuming that the connectivity features and the metabolic prior are estimated in the same spatial referential. The prior coefficients $\mathbf{w}_{\text{prior}}$ act as regularizers of a linear model on the functional connectivity. The model operates on the connectivity features $\mathbf{X}_{r_i}$ computed from (2). It integrates the prior within the penalization term yielding

$$\hat{\mathbf{w}}_{r_i} = \underset{\mathbf{w}\in\mathbb{R}^{N_v}}{\operatorname{argmin}} \|\mathbf{X}_{r_i}\mathbf{w} - \mathbf{y}\|_2^2 + \alpha\|\mathbf{w} - \lambda\mathbf{w}_{\text{prior}}\|_2^2, \tag{4}$$

where $\mathbf{w}_{\text{prior}}$ is the vector of weights that has already been learned. $\alpha > 0$ is a penalization parameter that controls the amount of shrinkage. We use the least-squares loss function so that the regularization can be easily integrated. By substituting $\mathbf{b} = \mathbf{w} - \lambda\mathbf{w}_{\text{prior}}$, one falls back to a classical ridge regression formulation. $\lambda > 0$ is a scaling parameter that adapts the prior to the actual setting. Here we assume that the discriminative weights of the metabolism and the functional connectivity are linearly correlated. When λ is zero, there is no effect of the prior, hence the weights estimated depends solely on connectivity features. Having λ too large imposes the rs-fMRI model to fully replicate the FDG-PET model, in effect underfitting the fMRI data. Model parameters λ and α are empirically estimated through a nested cross-validation on the training set. The resulting $\hat{\mathbf{w}}_{r_i} \in \mathbb{R}^{N_v}$ is the PET-informed coefficient vector of the classifier according to ROI r_i.

2.4 Stacking Connectivity-Based Predictions

We introduce here the stacking model to predict the diagnosis by learning from the predictions of the atlas ROIs. The unthresholded predictions from (4) of all ROIs are concatenated, which yields the following matrix: $\mathbf{S} \in \mathbb{R}^{N_s \times N_r}$

$$\mathbf{S} = \left(\mathbf{X}_{r_1} \hat{\mathbf{w}}_{r_1},\ \ldots\ , \mathbf{X}_{r_i} \hat{\mathbf{w}}_{r_i},\ \ldots\ , \mathbf{X}_{r_{N_r}} \hat{\mathbf{w}}_{r_{N_r}} \right), \tag{5}$$

where each column i represents the predictions of ROI r_i in all subjects. The subject class is learned from another classifier on the matrix $\mathbf{S}$, such as a logistic regression classifier:

$$\hat{\mathbf{w}}_s = \underset{\mathbf{w}\in\mathbb{R}^{N_r},\, c\in\mathbb{R}}{\operatorname{argmin}} \frac{C}{2}\mathbf{w}^T\mathbf{w} + \sum_{k=1}^{N_s} \log(\exp(-y_k(\mathbf{S}^{(\mathbf{k})}\mathbf{w} + c)) + 1). \tag{6}$$

where C controls the regularization, c is the intercept and $\hat{\mathbf{w}}_s$ are the estimated stacking model coefficients to define the final discriminative model. Since the stacking is actually compatible with any kind of classifier, non-linear classifiers (e.g. random forests) can be used as well. We discuss the choice of the classifier further by considering their impact on the model accuracies. We discuss also the impact of taking different sets of ROIs from different functional atlases.

3 Data and Experiments

3.1 ADNI Dataset

The 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, together with clinical and neuropsychological assessment, can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). For up-to-date information, see `www.adni-info.org`.

3.2 Subject Information

The subjects selected from the ADNI database in our study belong to two datasets (one dataset per modality). Let us emphasize that the PET dataset is different from the fMRI dataset; our goal is precisely to assess how metabolism

measurements can inform the connectivity without resorting to a multi-modal PET-fMRI prediction in each subject.

Table 2-a and Table 2-b summarize phenotype informations of the fMRI dataset and the PET dataset respectively. There are three possible values for the subject diagnosis: *i*) Alzheimer's Disease (AD) when the pathology has a histological confirmation by analyzing the cerebral spinal fluid collected from lumbar puncture; *ii*) Mild Cognitive Impairment (MCI) diagnosis gathers a spectrum from cognitive issues to prodromal stages of AD; this clinical group is generally determined through the mini-mental state examination score (MMSE) which relies on a test of functions like attention, calculation, recall, orientation, etc; *iii*) Cognitively Normal (CN) diagnosis is assigned if no evidence of a cognitive decline was detected during the MMSE (scores up to 30).

The PET dataset is used to build the metabolic prior. For this purpose, we select PET scans of 1371 subjects at baseline. Concerning the fMRI dataset, the available acquisitions in the database are quite limited, since resting-state protocol has been included only in the second phase of ADNI, 5 years after the beginning of the PET and the MRI data collection. Only 211 fMRI acquisitions at baseline are currently available. When available, we used repeated fMRI acquisitions from the same subjects, as we expect that more data will capture more variability and will improve the generalization of the learned models. A total of 694 fMRI scans from 211 subjects have been used in this study. However, learning on longitudinal data can yield biased and overfitted models if images from the same subject are included both in the training and the testing sets. To avoid such effects, we use a training/testing split at the subject level rather than at the acquisition level, while keeping the proportions between the clinical groups. The model training is done on all scans of all subjects included in the training phase. The model is tested on the remaining scans. The predictions are averaged for all scans of a given subject. A majority voting can be considered instead but we kept the averaging since we have a small number of scans per subject (five scans at most).

3.3 Image Acquisition and Preprocessing

The preprocessing of PET and fMRI scans was performed utilizing a combination of the Statistical Parametric Mapping (`SPM12`) software (`www.fil.ion.ucl.ac.uk/spm/`) for the classical preprocessing tasks, and the `Nilearn` Library (`nilearn.github.io`) [32] for the temporal preprocessing and the timeserie extraction. The acquisition and the preprocessing steps are detailed in the following subsections. In addition to the usual prepro-

Table 2: Demographic and neuropsychological characteristics.
The fMRI dataset is used for the classification, and the FDG-PET dataset is used for the prior computation. The subjects from the fMRI dataset are different than the ones from the FDG-PET dataset. We have more fMRI scans than subjects since we selected longitudinal fMRI acquisitions to have a better model generalization.

(a) - fMRI dataset

Diagnosis	N. Subj.	N. scans	Males	Age	MMSE
AD	36	117	44%	72.6 ± 6.8	21.8 ± 3.0
MCI	98	360	52%	72.2 ± 7.8	27.6 ± 2.0
CN	77	217	40%	74.2 ± 6.6	29.0 ± 1.1

(b) - FDG-PET dataset

Diagnosis	N. Subj.	Males	Age	MMSE
AD	273	58 %	74.9 ± 7.2	21.9 ± 3.4
MCI	665	58 %	73.4 ± 7.5	27.4 ± 2.2
CN	433	48 %	75.0 ± 7.1	29.1 ± 1.0

(AD: Alzheimer's Disease. MCI: Mild Cogntive Impairment. CN: Cognitively Normal. MMSE: Mini-Mental State Examination.)

cessing steps, the resulting images have been resampled in the same spatial referential (namely MNI152), so that images from both modalities can be mapped at the voxel level.

3.3.1 FDG-PET Images

FDG-PET images were acquired using a 18F-labelled fluoro-deoxyglucose (18F-FDG) radiotracer. SIEMENS, GE and PHILIPS PET scanners according to one of three standard protocols (30 - 60 minute dynamic, 30 - 60 minute static, 0 - 60 minute dynamic). The FDG-PET images used in this study were downloaded from the ADNI database website and were already preprocessed. Each PET image is coregistered to the first frame and the sequence is averaged into one frame. Then, the averaged images are standardized to have a uniform resolution and a voxel size of $3\times3\times3\,\mathrm{mm}^3$. FDG-PET image intensities are normalized to those of the pons so that the standard uptake

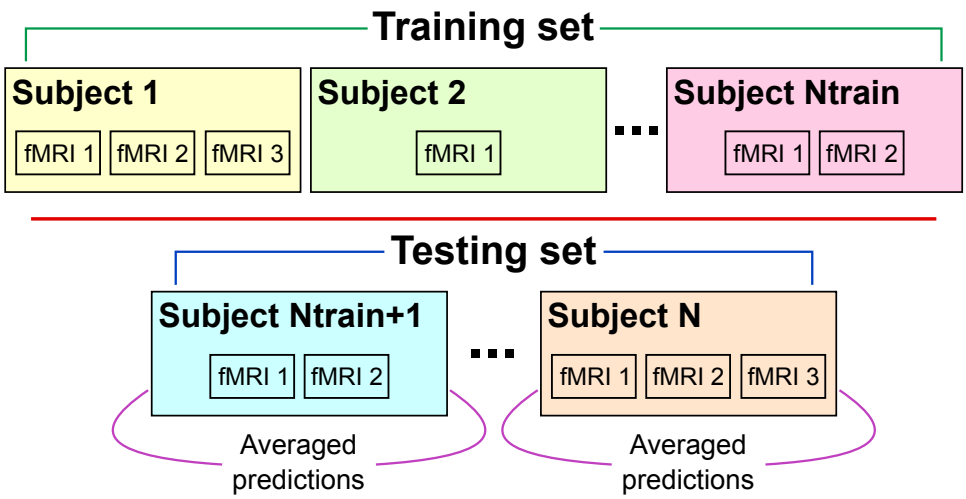


Figure 2: **Training/testing split scheme used on rs-fMRI data.** As the number of subjects is quite small, longitudinal acquisitions are included during the training phase. In order to avoid subject over-fitting, the split is done over the subjects and the prediction is obtained by averaging across acquisitions over each subject. The train/test split is iterated across 100 reshufflings.

values (SUV) of PET scans from different scanners can be compared.

3.3.2 rs-fMRI Images

All rs-fMRI data were acquired on 3.0 Tesla PHILIPS scanners at multiple sites. These scans are about 3 mm isotropic, with TE = 30 ms and TR = 3000 ms. The rs-fMRI sequence consists of 140 volumes (timepoints). Each frame has a shape of $64 \times 64 \times 48$ voxels. For the preprocessing of each acquisition, the first 5 frames are discarded. The remaining 135 frames of the acquisition are motion-corrected and coregistered to the corresponding T1 structural image. We use the DARTEL algorithm [33] to normalize the images; basically, a template is built at the group-level and each fMRI acquisition is non-linearly deformed to this template. The images are then spatially standardized and resampled into the MNI space, and smoothed using a Gaussian kernel of 6 mm FWHM. Timeseries are detrended for signal drift, and filtered with a $0.01 - 0.1$ Hz bandpass filter. In addition, some nuisance variables are removed like the global mean signal, the white matter and the cerebral spinal fluid signals.

3.4 Experiment Settings

The experiments are performed on the datasets described in 3.2. The trans-modal learning is performed and compared to learning on fMRI only, either with region-to-region or region-to-voxel connectivity. Then, atlas and classifier impact on the model are assessed by comparing different selections, in order to decide which atlas and classifier should be used. Finally, we experiment the stacking approach in a multi-modal classification by including features from complementary modalities.

Our aim is to study and predict conversion of MCI subjects to AD. However, the rs-fMRI protocol has been integrated only recently in the ADNI study (from ADNI-GO phase), and only 5 MCI converters in the database come with fMRI data. We thus consider a proxy by estimating the binary classification models to discriminate AD against MCI subjects. Classification models are assessed through the accuracy defined as the proportion (in %) of the correctly classified samples from all the samples of the test set. We use a stratified leave-k-out cross-validation scheme (k is 25% of the dataset size), where training and testing splits are randomized 100 times (see Fig. 2). Model hyper-parameters (α, λ) are set by a nested cross-validation in each randomization. It relies on a 4-fold cross-validation applied on the training set (75% of the dataset) to tune the prior scaling and the regularization parameters in (4). Classifier comparison is done by comparing the mean and the standard deviation of the accuracies, as well as a two-sided Wilcoxon significance test. All experiments were implemented in `Python`, using the `scikit-learn` library [34].

4 Results and Discussion

We present in this section experiments to demonstrate the effectiveness of the stacking approach and the metabolic prior integration to classify AD subjects from brain network connectivity on rs-fMRI data. We compare the accuracy of the classifiers when using the stacked region-to-voxel connectivity and the PET-informed model against classical approaches like region-to-region connectivity. We also study the impact of setting some parameters of the proposed approach, such as the brain atlas or the classification model. Then, we show the extensibility of the stacking framework to integrate multi-modal and non-image features that improve the overall accuracy of AD prediction.

4.1 AD Classification by Stacking rs-fMRI Connectivity Maps

In this experiment, we show how stacking region-to-voxel connectivity maps without the PET prior predicts AD by comparing it to the classification from region-to-region connectivity matrices. For this, we compute the learning curves for the two connectivity approaches, see Fig. 3. The learning curve represents variations of the classifier accuracy when increasing the number of samples during model training.

The first information from these curves is that the region-to-voxel stacking approach significantly outperforms the classification with the connectivity matrix between regions. This can be explained by the fact that the information is restricted in the region-to-region connectivity to the selected set of regions and does not capture the functional activity in the remaining brain structures. By contrast, the region-to-voxel approach with multiple ROIs is more exhaustive as it covers the whole brain, so that the classifier takes into account potentially more discriminative information. Previous experiments revealed also that stacking connectivities from several ROIs performs better than using a single ROI [27].

Second, we observe the benefit of including subject longitudinal acquisitions and averaging predictions of these acquisitions per subject rather than using single rs-fMRI at baseline. This improves slightly, yet consistently the accuracy, which was expected, since the training is done on three times more data.

Thirdly, we note that for all classification models considered, having more subjects in the training phase improves the classification accuracy, as all the curves display an increase with the number of samples. This suggests that using rs-fMRI as an imaging biomarker is relevant and could be even more accurate by enriching the classifier with more data. In this sense, the purpose of the integration of knowledge learned from the FDG-PET is to overcome the limitations imposed by the dataset size on the classifier accuracy. This will be discussed in section 4.3.

In summary, this experiment validates the stacking approach and the averaging of multiple datasets per subject. We study in the next section the impact of the classifiers in order to decide which model should be used for the stacking approach.

4.2 Which Classifier Should Be Used for Stacking?

We have presented in section 2.4 the stacking approach. On one hand, we propose to use a linear classifier that is well suited to the region-to-voxel

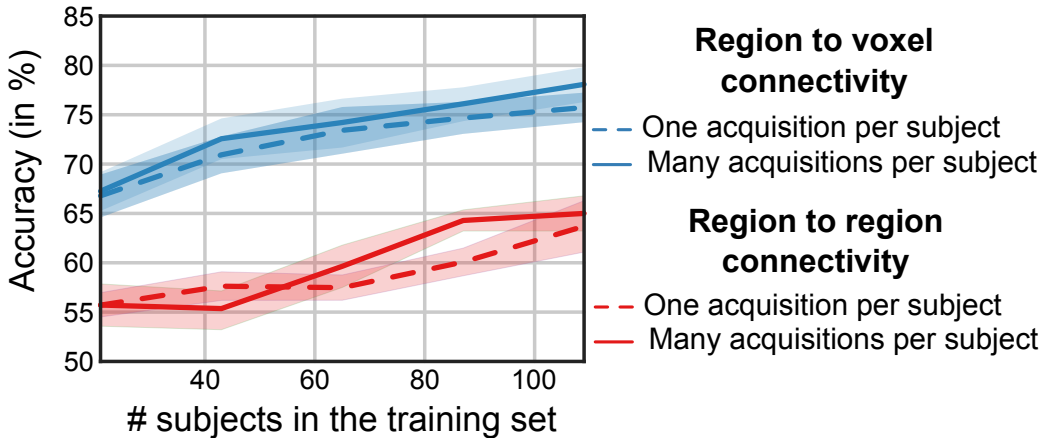


Figure 3: **Learning curves of AD/MCI classification based on functional connectivity features.** We observe that: *i*) The stacked region-to-voxel connectivity models are more accurate than the region-to-region connectivity models. *ii*) Using the subject longitudinal data in the training set gives better accuracy than using only one rs-fMRI per subject. *iii*) Including more samples in the training set improves the prediction accuracy in the testing set.

connectivity, as the model coefficients provide explicitly an estimate of the discriminative patterns between the clinical groups at the voxel level. On the other hand, there is no constraint on what classifier should be used for stacking the predictions. In order to select the best pair of classifiers, we compare several linear models for the connectivity predictions combined with the following three classifiers for the stacking step: logistic regression, ridge regression and random forests.

Table 3 reports the mean and the standard deviations of the AD/MCI classification accuracies with different classifiers tested. Each row is a linear model for the connectivity classification, while columns represent the stacking classifier. The main observation from the results is that all classifier combinations give almost the same mean accuracy around 77%. The choice of the linear model for region-to-voxel connectivity has no impact on accuracy. However, when stacking predictions, random forests classifier is more stable than logistic regression and ridge classifier. Standard deviations of accuracies over the runs is indeed significantly reduced. This is explained by the nature of random forests, which are an ensemble method where the prediction accuracy is stabilized by the internal averaging step; in other settings, this has also been reported to have a beneficial effect for prediction accuracy, even with no tuning [35].

Table 3: AD prediction accuracies according to several classifiers. While the mean accuracy is almost the same, stacking the connectivity predictions with Random Forests reduces the variance.

		Stacking		
		Ridge	Logistic Reg.	Random Forests
Connectivity	Ridge	77.2 ± 8.9	77.4 ± 8.4	77.3 ± **4.3**
	Logistic Reg.	76.9 ± 8.6	77.0 ± 8.6	76.8 ± **4.3**
	Linear SVC	77.2 ± 8.8	76.9 ± 8.3	76.8 ± **4.4**

4.3 Functional Connectivity Based Classification with Metabolic Prior Integration

We analyze in this experiment the effect of the metabolic prior integration on the classification of AD subjects from rs-fMRI. Box plots in Fig. 4 represent the variations of the classification accuracy over 100 randomizations when using different approaches (region-to-region connectivity, region-to-voxel connectivity and the PET informed region-to-voxel connectivity) with four brain atlases. The results show that the proposed classification scheme with the metabolic prior method outperformed pure functional connectivity methods. Accuracy improvements are about 5% without prior: 77%, with prior: 82%) and are statistically significant. This means that the discriminative patterns learned from metabolism data helped to recover a better model fron the functional connectivity data.

The prior is learned from FDG-PET data from the dataset described in Table 2-b. Basically, a linear classifier was estimated over this dataset, where metabolism values on whole brain voxels are the features of the model. We used a logistic regression for the prior estimation. Other linear classifiers have been tested and we did not find major differences in term of accuracy and model coefficient distribution. Resulting classifier weights are the discriminative coefficients between the metabolic activity of AD and MCI subjects. Fig. 5-a gives a plot of these coefficients after standardization and thresholding. The main discriminative structures are the typical cerebral regions, such as the posterior cingulate cortex (PCC), the precuneus, and parts of the parietal lobe. We find in these regions a metabolism that is significantly decreased for subjects with AD, which is in agreement with the findings in many studies, such as in [36]. These coefficients constitute the metabolic

prior map that constrains the classification of the functional connectivity from rs-fMRI. The model accuracy in Fig. 4 (last box plot on the right) validates the statistical power of the metabolic prior. The mean accuracy (88%) is similar to state-of-the-art results on FDG-PET data from ADNI database, although we cannot compare directly scores of studies with different subsets and features.

An interesting property of the stacking model is that we can recover the discriminative spatial map between AD and MCI subjects, since we perform a voxel-level brain analysis. The learned brain spatial models consist of averaged classifier weights across leave-k-out folds. Note that the resulting map is also an average of the coefficient maps of all ROIs, weighted by the importance of the ROI from the stacking with the random forests classifier. Fig. 5-b shows the normalized and thresholded coefficient map of the region-to-voxel connectivity based model without the FDG-PET prior. Although patterns are quite noisy, these regions describe some meaningful functional structures such as the default mode network, and parts of the parietal lobe. The impact of the FDG-PET prior is shown in Fig. 5-c, where we see that the metabolic prior overcame the limitations of connectivity-based discriminant patterns. We observe in particular patterns that are smoother than with fMRI only, e.g. the clearly outstanding default mode network. This finding is in agreement with AD studies that showed functional connectivity differences observed in rs-fMRI [12], but it is hard to obtain from small rs-fMRI datasets.

4.4 Impact of the Brain Atlas on the Classification

We assess the impact of choosing a given ROIs-defining atlas on classification accuracy. For this, we compare four atlases through AD/MCI classification. These atlases have different numbers of regions and have different nature (functional, structural, histological). The first functional atlas (MSDL) has been proposed in [37], it contains 39 ROIs learned from rs-fMRI data by a group-level dictionary-learning decomposition. The second atlas (Mayo Clinic) comprises 68 ROIs extracted from a functional dataset and proposed in [38]. It was constructed on 892 subjects by an independent component analysis. It yields a detailed decomposition of the default mode network, and was successfully used to characterize differences between AD and cognitively normal subjects. The third and fourth atlases are respectively structural and histological atlases. The Harvard-Oxford (96 cortical ROIs) and the Jülich (120 ROIs) atlases were obtained from FMRIB Software Library (`FSL` [39]) [1]

[1] For more details about these two atlases see `fsl.fmrib.ox.ac.uk/fsl/fslwiki/Atlases`.

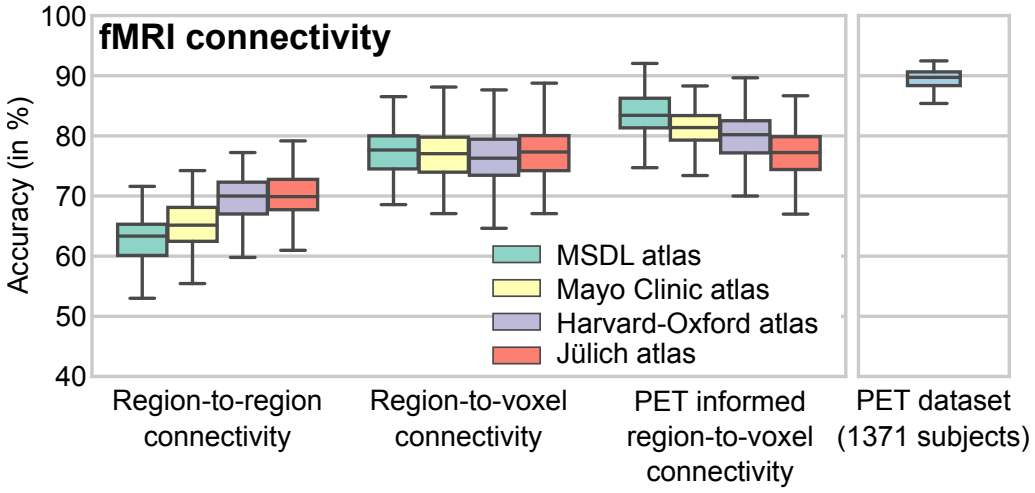
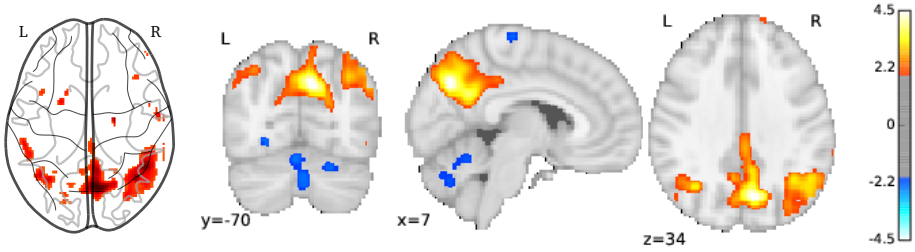


Figure 4: **Impact of the transmodal connectivity on AD/MCI classification.** The *stacking + prior* approach is more accurate than *stacking without prior* or using region-to-region connectivity. Four different atlases are compared here : *i*) MSDL functional atlas [37] (39 regions). *ii*) Mayo Clinic functional atlas [38] (68 regions). *iii*) Harvard-Oxford anatomical atlas (96 regions). *iv*) Jülich Anatomical atlas (120 regions). The inter-atlas comparisons show that there are no major differences in the accuracy.

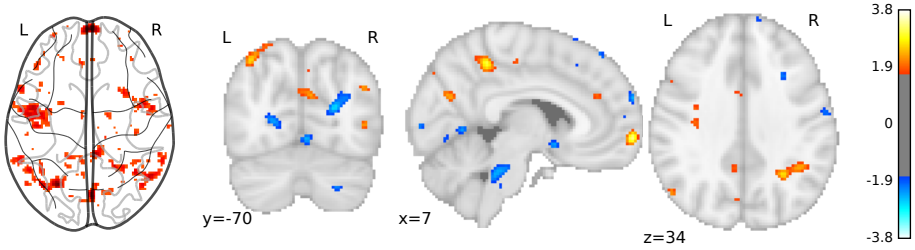
The accuracy obtained with the competing atlases is shown in Fig. 4, where the results are grouped according to the classification approach. The atlas choice has no impact on the region-to-voxel connectivity without the metabolic prior, as classification accuracies are almost similar. When integrating the metabolic prior, we observe that functional atlases are more accurate than structural atlases, although the differences are not statistically significant. This means that the metabolic prior may not fit well with connectivity maps of non functional regions, and that adding non relevant regions does not improve stacking accuracy. Regarding the region-to-region connectivity, increasing the number of ROIs gives better accuracy. This effect is explained by the fact that taking more ROIs will include more brain structures that may have discriminative information. Overall, the complexity of the model used yields different bias/variance compromises that condition the choice of the atlas.

4.5 Extension to Multi-modal Stacking

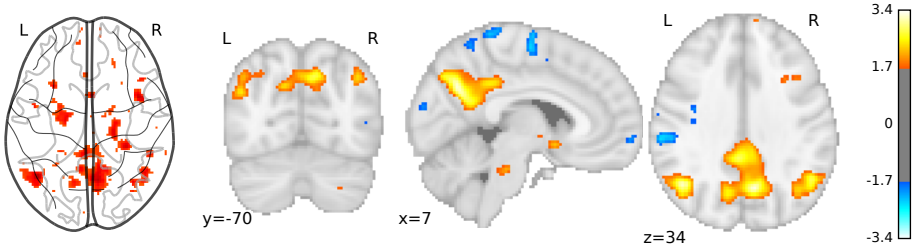
Here we demonstrate how the stacking framework can be extended to multimodal data. This experiment also evaluates the relevance of rs-fMRI con-



(a) PET prior map.



(b) rs-fMRI connectivity stacking map.



(c) PET informed rs-fMRI connectivity stacking map.

Figure 5: **Spatial maps of AD/MCI classification coefficients.** The parts of the default-mode network (involved in AD characterization) are better recovered from the prior-informed stacking model than the stacking model without prior, since it has been imposed during the regularization. The final stacking map is an aggregate of atlas ROIs weighted by stacking model coefficients. The maps are normalized and thresholded.

nectivity informed by metabolic prior in a multimodal classification context of AD.

Many works have shown that combining complementary information about anatomical features from structural MRI, functional features from FDG-PET, and biological features from cerbrospinal fluid (CSF) leads to a more accurate prediction of AD [40], [41]. In our setting, we combine PET-informed fMRI connectivities from fMRI dataset at baseline with features from anatomical MRI and CSF. These features were collected from processed

and quality-checked data uploaded in ADNI database. For the CSF, three biomarkers measurements (A$\beta_{1\text{-}42}$, t-tau, p-tau$_{181}$) were extracted from an analysis done at the University of Pennsylvania [42]. Regarding the anatomical MRI, we selected sixteen volumetric features of segmented hippocampus from a processing using `FreeSurfer` software performed at the university of California San Francisco. Thus, each subject is represented by a feature vector composed of the three modalities. We follow the same cross-validation scheme as presented in 3.4, by using a stratified leave-25%-out randomized over 100 runs.

Fig. 6 summarizes accuracies of the AD/MCI classification on each single set separately (PET-informed connectivity, CSF biomarkers, hippocampus volumetry), then by combining modalities either by concatenating or by applying multiple kernel learning, and by stacking predictions from each modality classifier. The main message from these results is that the functional connectivity is a biomarker that brings a complementary information to reference biomarkers like CSF proteins and hippocampus volumetry, since combining these features yields better classification accuracy than using a single modality, which is in agreement with recent studies on predicting AD with rs-fMRI [43]. This suggests that fMRI could be used as a noninvasive alternative of the FDG-PET as a functional biomarker of AD. Moreover, stacking predictions of each modality is also a valid way to combine multimodal data as the accuracies (87%) are better than the concatenation (84%). Existing studies proposed to combine heterogeneous features from ADNI database for AD prediction, such as using multiple kernel learning [44], or random forests classifier [45]. We cannot compare directly our accuracy scores to the ones from the papers cited above since datasets are different. Hence we applied MKL on our dataset, by using an implementation of the MKL proposed in [28] on linear sub-kernels. We found that accuracies are in the same range as stacking (MKL accuracy: 86%, stacking accuracy: 87%). The benefit of using the stacking approach is that it enables explicit interpretation of heterogeneous modalities and can leverage variable importance computed by random forests. Indeed, each variable for the stacking represents a prediction from a distinct modality, so that the discriminative maps are recovered for example by returning to the first-level classifiers. It is also flexible to different classification schemes, since no assumptions are made on the classifier.

5 Conclusion

We have introduced transmodal learning in neuroimaging-based diagnosis, by enhancing classification of AD subjects from rs-fMRI with a data-driven

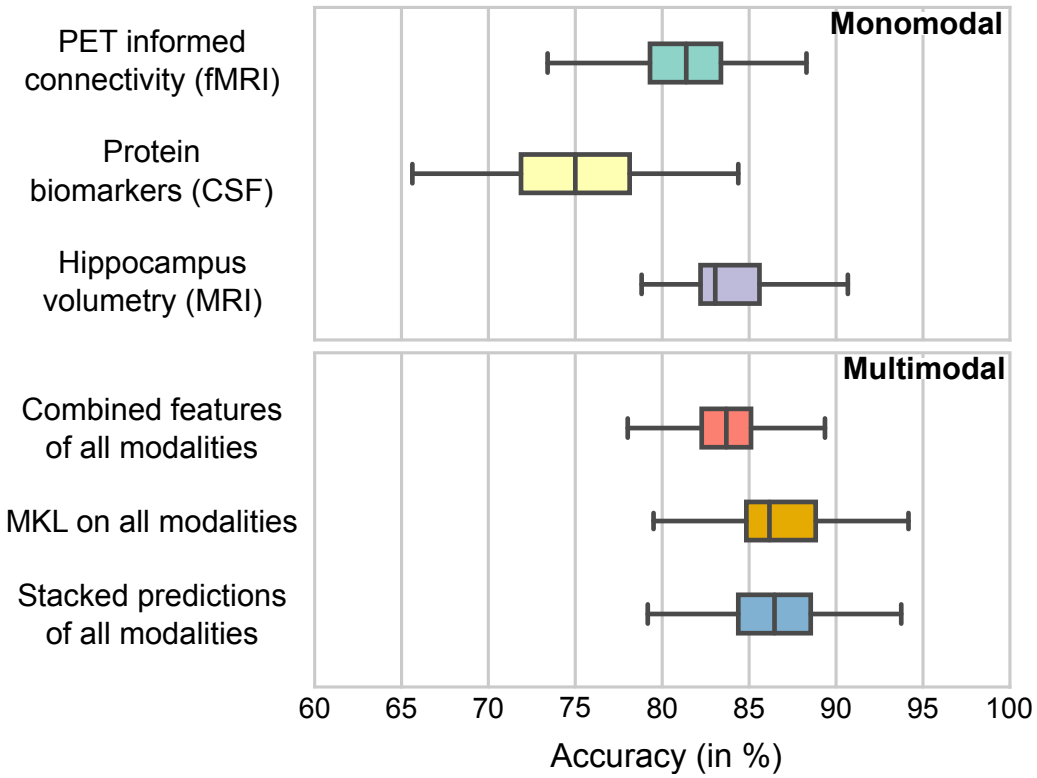


Figure 6: **Multimodal AD/MCI classification accuracy with the stacking framework.** The modalities used are: PET-informed and stacked connectivity from fMRI, cerebrospinal fluid (CSF) biomarkers ($A\beta_{1\text{-}42}$, t-tau, p-tau_{181}) and hippocampus volumetric features extracted from anatomical MRI. The results show that the functional information from connectivity adds accuracy to the hippocampus volumetry which is considered as a reference biomarker for AD. The stacking framework can be easily extended to a multi-modal framework. The stacking accuracies are comparable to multiple kernel learning and are better than feature concatenation.

FDG-PET prior learned from a distinct and larger cohort. Specifically, we call transmodal an approach using an imaging modality to inform another imaging modality on different subjects. Rather than combining multi-modal images, transmodal imaging does not rely on having images of both modalities from the same subjects. From an application perspective it enriches an imaging modality without requiring additional measurements.

For our application, experimental results confirm that metabolic activity of brain structures measured on FDG-PET images is linked to connectivity measured with resting-state fMRI. The transmodal approach improves

biomarker performance on a noninvasive modality (rs-fMRI) using an invasive but more sensitive modality (FDG-PET). Hence the metabolic prior can be used for further rs-fMRI acquisitions dataset without having recourse to FDG-PET acquisitions and to a multimodal analysis. Such an approach overcomes the limitations of fMRI (non-quantitative, low SNR), yielding accurate predictions of AD based on functional connectivity.

We find that characterizing functional networks on rs-fMRI with multiple region-to-voxel connectivity maps gives a more accurate discrimination of AD than region-to-region connectivity. We mitigate the difficult choice of seed ROI by combining maps derived from an atlas of ROIs that covers sufficiently the cortical surface. Then a stacking step combines the set of maps and improves prediction accuracy. We find that using random forests for stacking gives predictions that are not only accurate, but also stable. In addition, stacking is interpretable as it recovers discriminative maps related to the state-of-the-art on AD. Finally, stacking is interesting as it can be extended to multimodal data. Indeed, fMRI is used here as an indicator of functional activity, but anatomical imaging and non-imaging markers capture other complementary aspects of disease progression. Combining these in the stacking step improves the accuracy of the predictive model.

Future work will tackle the high dimensionality of voxel-based (whole brain) connectivity. Dimensionality reduction with clustering methods might be useful to solve this problem [46, 47]. Also, the metabolic prior may not be suitable for some ROIs that are not related to the main resting-state networks. Regarding the dataset size, we observe that learning on multiple scans per subject improves the model. In this sense, existing datasets are a very useful resource to investigate more difficult questions such as the prognosis of conversion from MCI to AD. A truly longitudinal approach, making use of time-evolution data for each subject, would probably yield even better results, and provide new insights on the pathology progression and its impact on the resting-state networks.

Acknowledgements

This work is supported by the Lidex PIM project funded by the IDEX Paris-Saclay, ANR-11-IDEX-003-02. We also acknowledge funding from the NiConnect project (ANR-11-BINF-0004_NiConnect).

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 Ag-

ing, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's 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.

References

[1] R. Cuingnet, E. Gerardin *et al.*, "Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database." *NeuroImage*, vol. 56, 2011.

[2] M. R. Sabuncu, R. S. Desikan *et al.*, "The dynamics of cortical and hippocampal atrophy in Alzheimer disease." *Archives of neurology*, vol. 68, 2011.

[3] F. Prados, M. Jorge *et al.*, "Measuring brain atrophy with a generalised formulation of the boundary shift integral," *Neurobiology of Aging*, vol. 36, 2015.

[4] P. T. Trzepacz, P. Yu *et al.*, "Comparison of neuroimaging modalities for the prediction of conversion from mild cognitive impairment to alzheimer's dementia," *Neurobiology of Aging*, vol. 35, 2014.

[5] N. Villain, G. Chételat *et al.*, "Regional dynamics of amyloid-β deposition in healthy elderly, mild cognitive impairment and Alzheimer's disease:A voxelwise PiB-PET longitudinal study," *Brain*, vol. 135, 2012.

[6] K. R. Gray, R. Wolz, R. A. Heckemann, P. Aljabar, A. Hammers, and D. Rueckert, "Multi-region analysis of longitudinal FDG-PET for the classification of Alzheimer's disease." *NeuroImage*, vol. 60, 2012.

[7] P.-J. Toussaint, V. Perlbarg *et al.*, "Resting state FDG-PET functional connectivity as an early biomarker of Alzheimer's disease using conjoint univariate and independent component." *NeuroImage*, vol. 63, 2012.

[8] F. X. Castellanos, A. Di Martino, R. C. Craddock, A. D. Mehta, and M. P. Milham, "Clinical applications of the functional connectome." *NeuroImage*, vol. 80, pp. 527–40, 2013.

[9] E. L. Dennis and P. M. Thompson, "Functional brain connectivity using fMRI in aging and Alzheimer's disease." *Neuropsychology review*, vol. 24, no. 1, pp. 49–62, Mar. 2014.

[10] J. Richiardi, H. Eryilmaz, S. Schwartz, P. Vuilleumier, and D. Van De Ville, "Decoding brain states from fMRI connectivity graphs." *NeuroImage*, vol. 56, no. 2, pp. 616–26, 2010.

[11] K. Wang, M. Liang, L. Wang, L. Tian, X. Zhang, K. Li, and T. Jiang, "Altered functional connectivity in early Alzheimer's disease: A resting-state fmri study," *Human brain mapping*, vol. 28, no. 10, 2007.

[12] M. D. Greicius, G. Srivastava *et al.*, "Default-mode network activity distinguishes Alzheimer's disease from healthy aging:evidence from functional MRI," *PNAS*, vol. 101, 2004.

[13] W. Koch, S. Teipel, S. Mueller *et al.*, "Diagnostic power of default mode network resting state fMRI in the detection of Alzheimer's disease," *Neurobiology of Aging*, vol. 33, no. 3, 2012.

[14] X. Jiang, X. Zhang, and D. Zhu, "Intrinsic functional component analysis via sparse representation on Alzheimer's disease neuroimaging initiative database." *Brain connectivity*, vol. 4, no. 8, pp. 575–86, Oct. 2014.

[15] A. S. Choe, C. K. Jones *et al.*, "Reproducibility and temporal structure in weekly resting-state fMRI over a period of 3.5 years," *PLOS ONE*, vol. 10, 2015.

[16] D. Tomasi and N. D. Volkow, "Resting functional connectivity of language networks: characterization and reproducibility," *Molecular Psychiatry*, vol. 17, no. 8, pp. 841–854, jan 2012.

[17] V. Riedl, K. Bienkowska *et al.*, "Local activity determines functional connectivity in the resting human brain: a simultaneous FDG-PET/fMRI study." *The Journal of neuroscience*, vol. 34, 2014.

[18] S. Passow, K. Specht *et al.*, "Default-mode network functional connectivity is closely related to metabolic activity," *Human Brain Mapping*, vol. 2038, 2015.

[19] M. Aiello, E. Salvatore, *et al.*, "Relationship between simultaneously acquired resting-state regional cerebral glucose metabolism and functional MRI:A PET/MR hybrid scanner study." *NeuroImage*, vol. 113, 2015.

[20] D. A. Gusnard and M. E. Raichle, "Searching for a baseline:functional imaging and the resting human brain." *Nature reviews.*, vol. 2, 2001.

[21] B. Ng, G. Varoquaux *et al.*, "A novel sparse graphical approach for multimodal brain connectivity inference," in *MICCAI*, 2012.

[22] X. Jiang, T. Zhang, Q. Zhao, J. Lu, L. Guo, and T. Liu, "Fiber connection pattern-guided structured sparse representation of whole-brain fMRI signals for functional network inference," in *MICCAI*, 2015.

[23] A. Venkataraman, Y. Rathi, M. Kubicki, C. Westin, and P. Golland, "Joint modeling of anatomical and functional connectivity for population studies," *IEEE Transactions on Medical Imaging*, vol. 31, 2012.

[24] B. Ng, G. Varoquaux, J. B. Poline, and B. Thirion, "Implications of inconsistencies between fmri and dmri on multimodal connectivity estimation," in *MICCAI*, 2013.

[25] S. J. Pan and Q. Yang, "A survey on transfer learning," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 22, p. 1345, 2010.

[26] B. Cheng, M. Liu, D. Zhang, B. C. Munsell, and D. Shen, "Domain transfer learning for MCI conversion prediction," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 7, pp. 1805–1817, jul 2015.

[27] M. Rahim, B. Thirion, A. Abraham, M. Eickenberg, E. Dohmatob, C. Comtat, and G. Varoquaux, "Integrating multimodal priors in predictive models for the functional characterization of alzheimer's disease," in *MICCAI*, 2015.

[28] A. Rakotomamonjy, F. Bach, S. Canu, and Y. Grandvalet, "Simplemkl," *Journal of Machine Learning Research*, vol. 9, pp. 2491–2521, 2008.

[29] D. M. Cole, S. M. Smith, and C. F. Beckmann, "Advances and pitfalls in the analysis and interpretation of resting-state FMRI data," *Frontiers in systems neuroscience*, p. 8, 2010.

[30] A. Zalesky, A. Fornito, and E. Bullmore, "On the use of correlation as a measure of network connectivity," *NeuroImage*, vol. 60, no. 4, 2012.

[31] G. Varoquaux and R. C. Craddock, "Learning and comparing functional connectomes across subjects," *NeuroImage*, vol. 80, pp. 405–415, 2013.

[32] A. Abraham, F. Pedregosa, M. Eickenberg, P. Gervais, A. Mueller, J. Kossaifi, A. Gramfort, B. Thirion, and G. Varoquaux, "Machine learning for neuroimaging with scikit-learn." *Frontiers in neuroinformatics*, vol. 8, no. February, p. 14, 2014.

[33] J. Ashburner, "A fast diffeomorphic image registration algorithm," *NeuroImage*, vol. 38, no. 1, pp. 95–113, 2007.

[34] F. Pedregosa, G. Varoquaux *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, 2011.

[35] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, 2001.

[36] L. Mosconi, W. H. Tsui, K. Herholz, A. Pupi *et al.*, "Multicenter standardized 18F-FDG PET diagnosis of mild cognitive impairment, Alzheimer's disease, and other dementias." *Journal of nuclear medicine*, vol. 49, no. 3, 2008.

[37] G. Varoquaux, A. Gramfort, F. Pedregosa, V. Michel, and B. Thirion, "Multi-subject dictionary learning to segment an atlas of brain spontaneous activity," in *Information Processing in Medical Imaging*, 2011.

[38] D. T. Jones, P. Vemuri, M. C. Murphy, J. L. Gunter, M. L. Senjem, M. M. Machulda, S. A. Przybelski, B. E. Gregg, K. Kantarci, D. S. Knopman, B. F. Boeve, R. C. Petersen, and C. R. Jack, "Non-stationarity in the "resting brain's" modular architecture," *PloS one*, vol. 7, 2012.

[39] S. M. Smith, M. Jenkinson, M. W. Woolrich, C. F. Beckmann, T. E. Behrens, H. Johansen-Berg, P. R. Bannister, M. De Luca, I. Drobnjak, D. E. Flitney *et al.*, "Advances in functional and structural mr image analysis and implementation as fsl," *Neuroimage*, vol. 23, 2004.

[40] K. B. Walhovd, A. M. Fjell, J. Brewer *et al.*, "Combining MR imaging, positron-emission tomography, and CSF biomarkers in the diagnosis and prognosis of Alzheimer disease." *Am. j. of neuroradiology*, vol. 31, 2010.

[41] C. Hinrichs, V. Singh, G. Xu, and S. C. Johnson, "Predictive markers for AD in a multi-modality framework: An analysis of MCI progression in the ADNI population," *NeuroImage*, vol. 55, no. 2, pp. 574–589, 2011.

[42] L. M. Shaw, H. Vanderstichele, M. Knapik-Czajka, M. Figurski *et al.*, "Qualification of the analytical and clinical performance of CSF biomarker analyses in ADNI," *Acta Neuropathologica*, vol. 121, 2011.

[43] V. Kebets, J. Richiardi, M. Van Assche *et al.*, "Predicting pure amnestic mild cognitive impairment conversion to alzheimer's disease using joint modeling of imaging and clinical data," in *Pattern Recognition in NeuroImaging (PRNI)*, 2015.

[44] D. Zhang, Y. Wang, L. Zhou, H. Yuan, and D. Shen, "Multimodal classification of Alzheimer's disease and mild cognitive impairment." *NeuroImage*, vol. 55, no. 3, pp. 856–67, 2011.

[45] K. R. Gray, P. Aljabar, R. a. Heckemann, A. Hammers, and D. Rueckert, "Random forest-based similarity measures for multi-modal classification of Alzheimer's disease," *NeuroImage*, vol. 65, pp. 167–175, 2013.

[46] B. Thirion, G. Varoquaux, E. Dohmatob *et al.*, "Which fMRI clustering gives good brain parcellations?" *Frontiers in neuroscience*, vol. 8, 2014.

[47] V. Michel, A. Gramfort, G. Varoquaux, E. Eger, C. Keribin, and B. Thirion, "A supervised clustering approach for fmri-based inference of brain states," *Pattern Recognition*, vol. 45, p. 2041, 2012.