Towards a better predictive model from rest fMRI: benchmarks across multiple phenotypes

Kamalaker Dadi[1*], **Darya Chyzhyk**[1], **Alexandre Abraham**[1], **Mehdi Rahim**[1], **Bertrand Thirion**[1], **Gael Varoquaux**[1]

[1]Parietal team, Neurospin – I[2]BM – CEA, INRIA, Paris-Saclay, France
*kamalaker-reddy.dadi@inria.fr

1. Introduction

Psychiatry and psychology are based on assessing individuals traits, characterized through behavioral testing and questionnaires. Imaging of brain activity raises the hope of measuring the physiological differences that underlie these psychological variations[1]. In[2], we have introduced an automated pipeline capable of learning this link across individuals using large cohorts of functional magnetic resonance images acquired during rest (Rest fMRI). We present the outline of this pipeline and how we used it to draw best practices from its application on various problems.

2. A connectome classification pipeline

We applied the pipeline on five datasets to i) determine the steps to obtain the best prediction, and ii) predict phenotypic information with good accuracy.

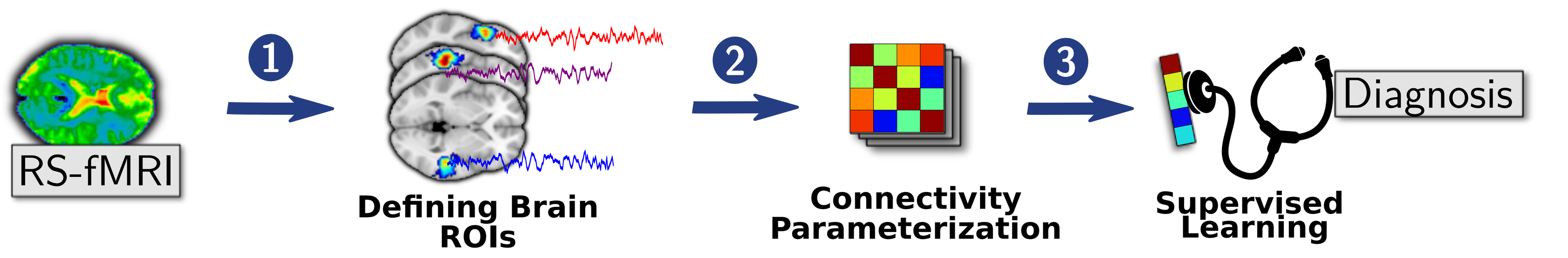


Figure 1: **Our pipeline consists of three main steps**: defines regions from rest fMRI, builds connectomes from time series signals extracted from these regions of interests, and compares connectomes across subjects using machine learning.

3. Rest fMRI datasets

Dataset	Subjects	Clinical question
COBRE	65/77	Schizophrenia vs Control
ADNI	40/96	AD vs MCI
ADNIDOD	89/78	PTSD vs Control
ACPI	62/64	Marijuana use vs Control
ABIDE	402/464	Autism vs Control

Table 1: Description of the five rs-fMRI datasets used. AD - Alzheimer's Disease, MCI - Mild Cognitive Impairment, PTSD - Post Traumatic Stress Disorder.

4. Pipelining choices

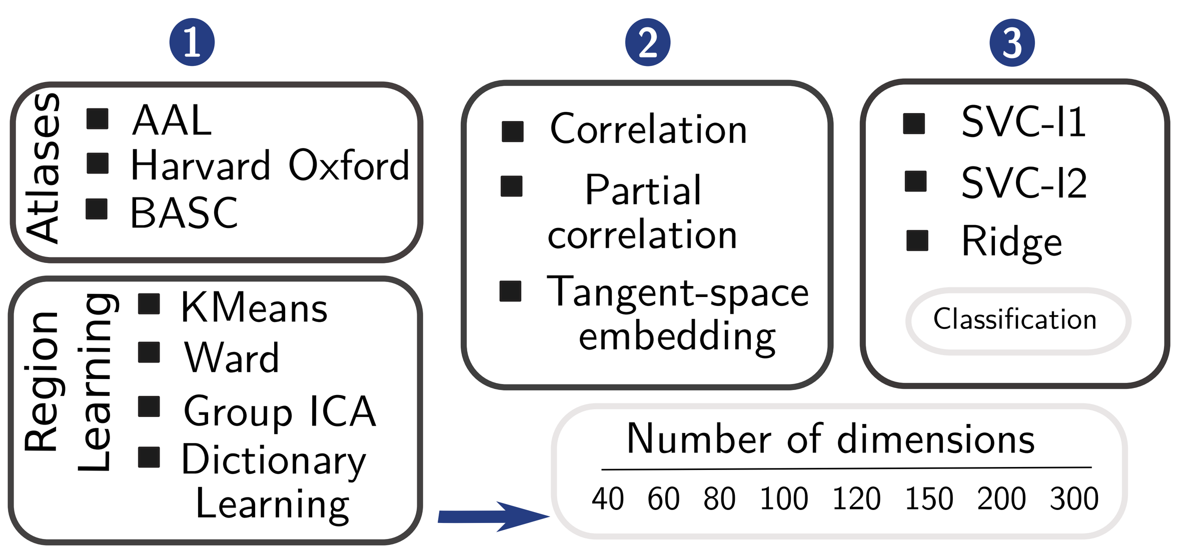
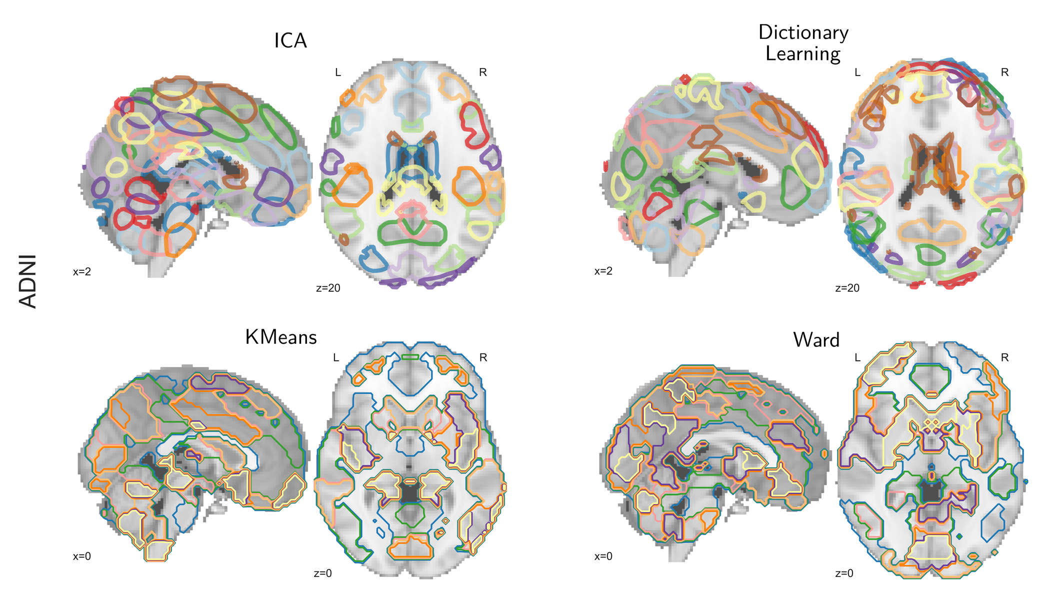


Figure 2: **1 – Defining Brain ROIs** can be done using three pre-defined atlases: two anatomical - Automated Anatomical Labeling (AAL), Harvard Oxford and one functional - Bootstrap Analysis of Stable Clusters (BASC). Four region learning methods: two Linear decomposition models - Group Independent Component Analysis (Group ICA), Online Dictionary Learning and two Clustering models - KMeans, Agglomerative with Ward criterion. **2 – Parameterizing functional connectivity between ROIs**: correlation, partial correlation or tangent space embedding **3 – Supervised learning**: Classifiers, a classification model is built to predict groups with two linear classifiers, SVC (ℓ_1 or ℓ_2 penalization) and Ridge. Definition of brain regions (below).



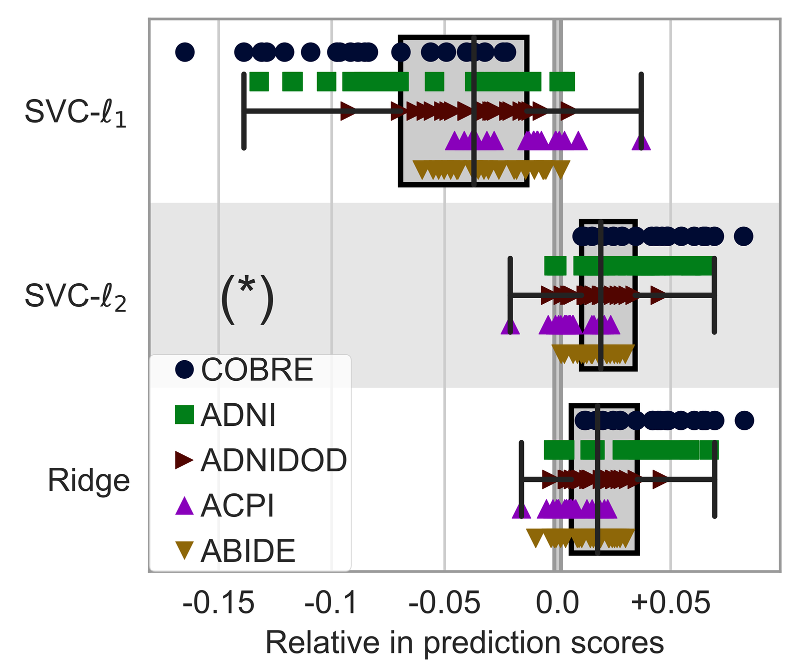
5. Results

5.1 Prediction scores in AUC

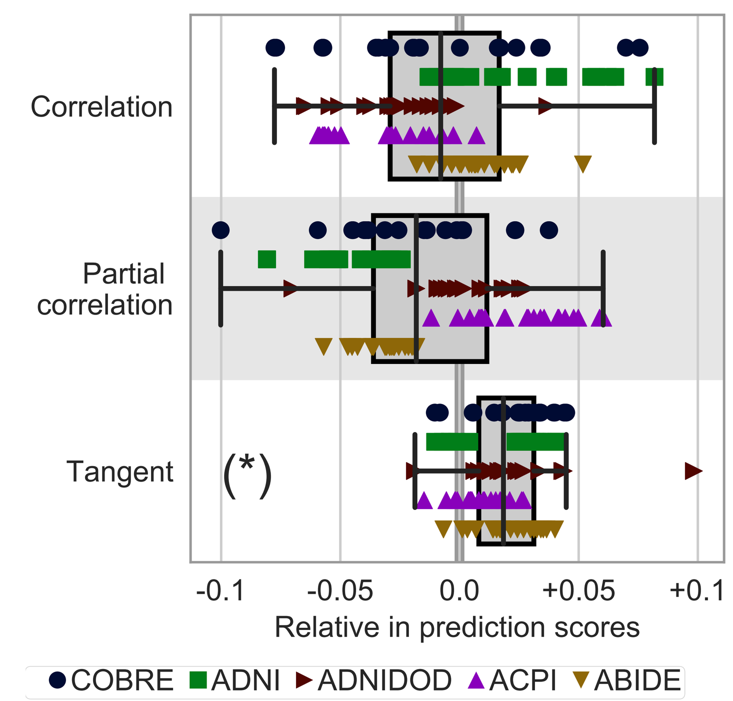
Accuracy	COBRE	ADNI	ADNIDOD	ACPI	ABIDE
Median	87.1%	76.2%	67.1%	56.4%	70.2%
5^{th} Percentile	74.3%	62.3%	53.8%	41.4%	64.9%
95^{th} Percentile	94%	89.1%	77.2%	68.7%	73.8%

Table 2: **Median, 5th percentile and 95th percentile of accuracy scores in AUC over cross-validation folds ($n = 100$) across multi rs-fMRI datasets.** Functional brain atlases learned using Online Dictionary Learning[3] with 60 resting state networks and spliting networks to regions are used in the prediction pipeline with Tangent based connectivity matrix parameterization[4] and Support Vector Classifier with ℓ_2 penalization (SVC-ℓ_2).

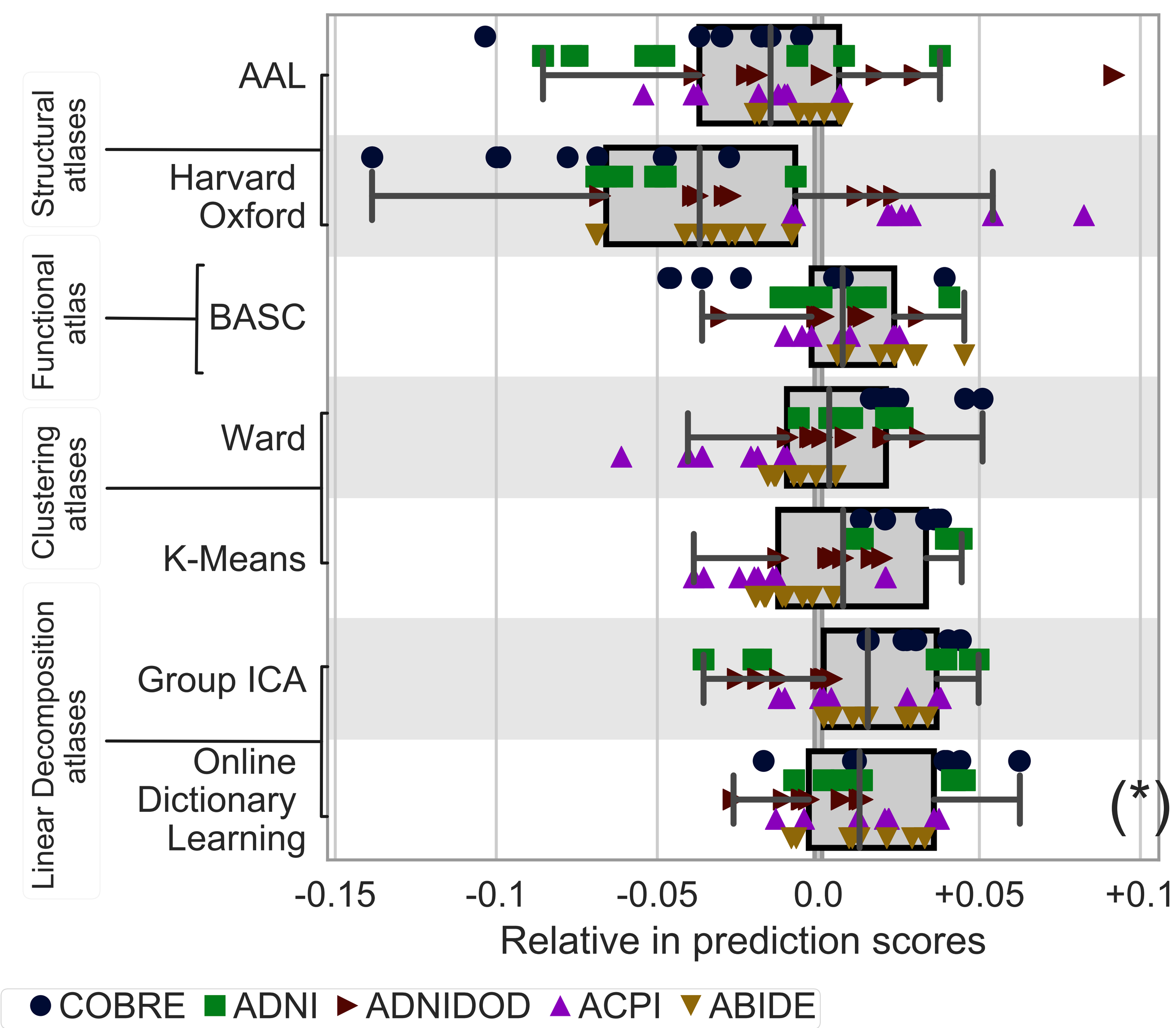
5.2 Choice of classifier, connectivity and brain atlas



(a) classifier



(b) connectivity parameterization



(c) brain atlases with optimal choices in dimensionality – BASC (122 networks to regions) – Ward (120) – K-Means (120 parcellations) – GroupICA (80 networks to regions) – DictLearn (60 networks to regions)

Figure 3: The **impact of choices** in functional connectivity prediction pipeline over diverse tasks. (*) indicates the optimal option.

6. Optimal choices

- Systematic exploration of choices at each step
- Definition of Brain ROIs - Linear Decomposition models
- Connectivity Parameterization - Tangent Space Embedding
- Classification with Support Vector Classifier - ℓ_2 penalization.

References

1. Karla, M. Multimodal population brain imaging in the uk biobank prospective epidemiological study. *Nature Neuroscience* (2016). 2. Abraham, A. Deriving reproducible biomarkers from multi-site resting-state data: An autism-based example. *Neuroimage* (2016). 3. Mensch, A. Compressed online dictionary learning for fast resting-state fmri decomposition. *ISBI* (2016). 4. Varoquaux, G. Detection of brain functional-connectivity difference in post-stroke patients using group-level covariance modeling. *MICCAI* (2010).

PARIETAL NeuroSpin Inria

Acknowledgments:
ANR-11-BINF-0004_NiConnect and CATI project.