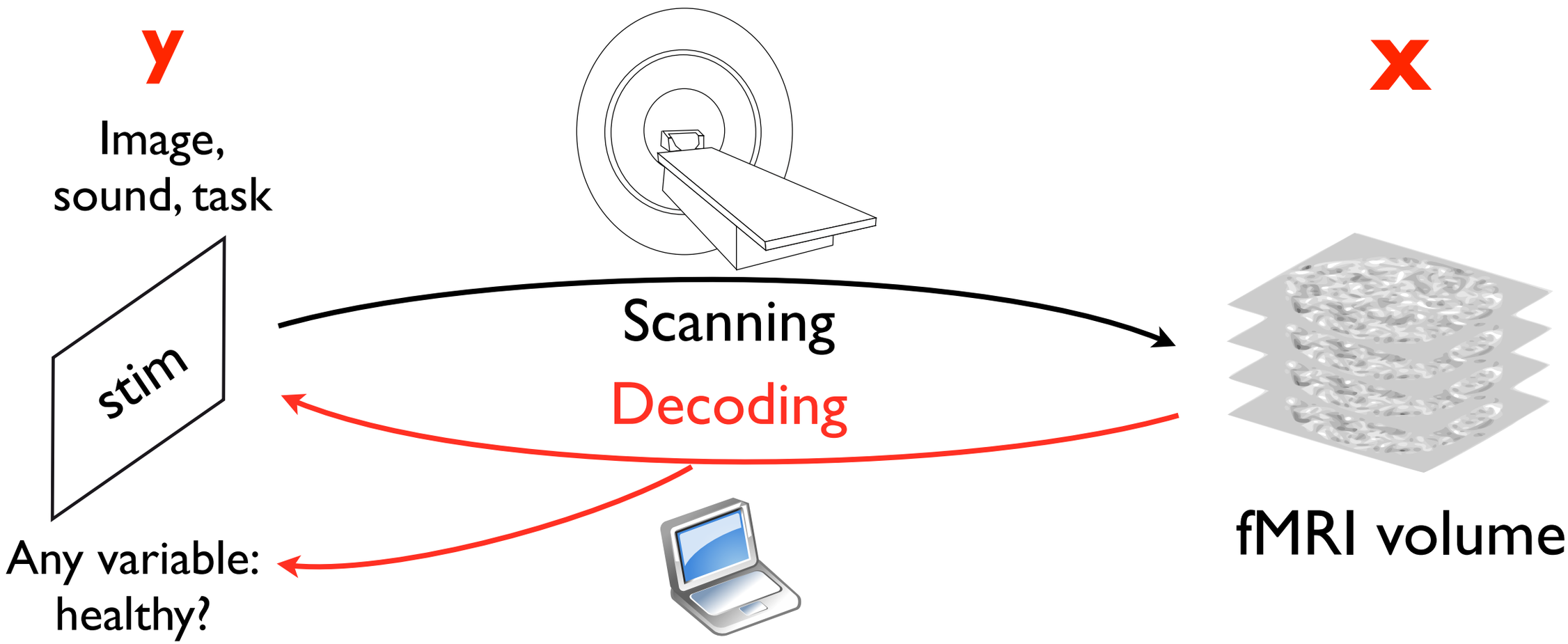
Chaire ANR IA: "BrAIN" Bridging Artificial Intelligence and Neuroscience

Alexandre Gramfort
alexandre.gramfort@inria.fr

INRIA, Université Paris-Saclay
CEA Neurospin

Sept. 2020

Supervised learning with fMRI



Objective: Predict y given X or learn a function

$$f : X \rightarrow y$$

Precision medicine / Biomarkers

IMPAC

IMaging-PsychiAtry Challenge: predicting autism

A data challenge on Autism Spectrum Disorder detection

Deadline: July 1, 2018 - 8 pm (UTC)

Days | Hours | Minutes | Seconds

https://paris-saclay-cds.github.io/autism_challenge/

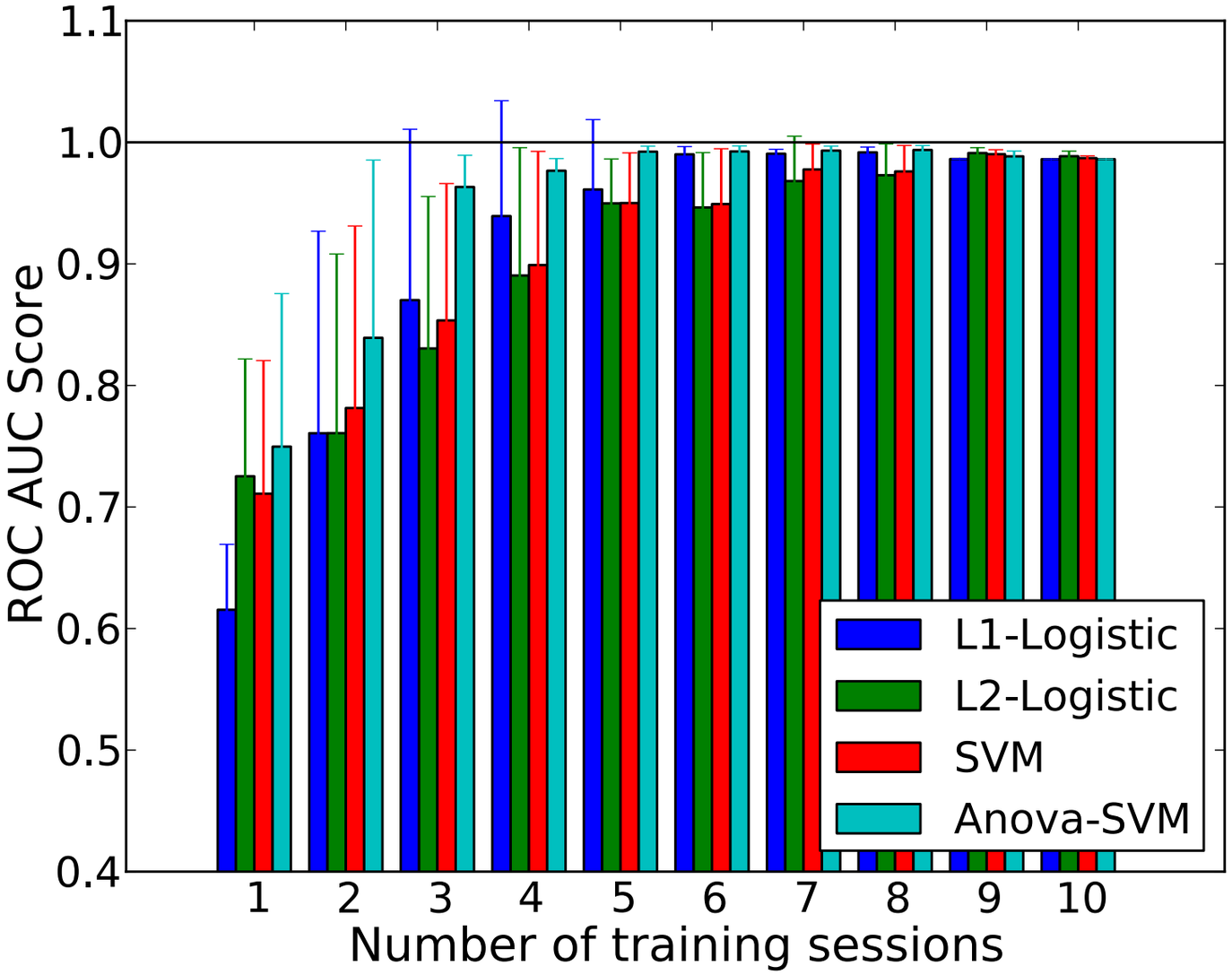
Paris-Saclay Center for Data Science

Hôpital universitaire Robert-Debré

ASSISTANCE PUBLIQUE HÔPITAUX DE PARIS

Institut Pasteur

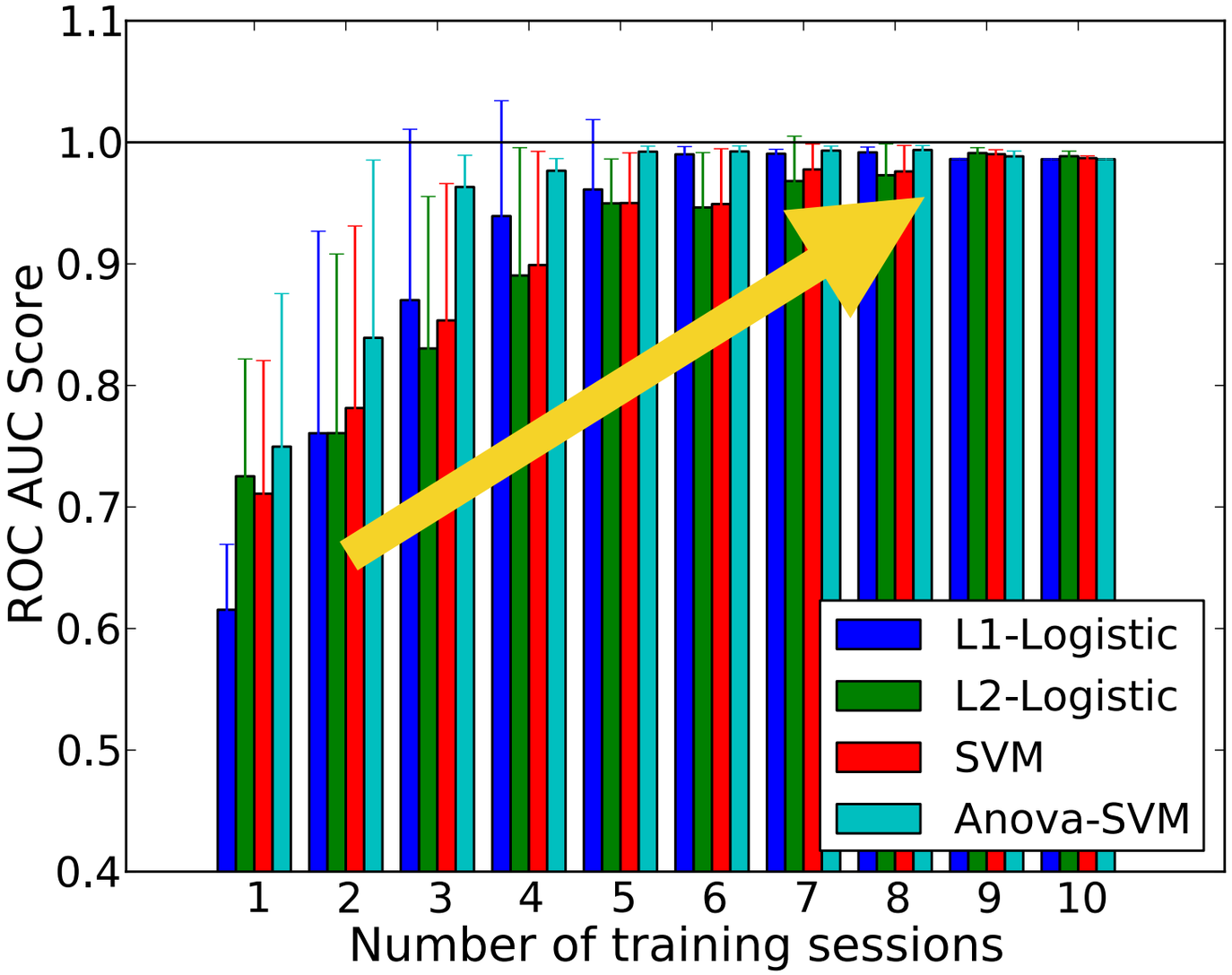
Why more data is better?



- 5 subjects
- 12 sessions (more than 1000 scans)
- Binary classification (face vs. house)
- Test of 2 left-out sessions

Data from *[Haxby et al. 2001]*
Figure from *[Gramfort et al. 2011]*

Why more data is better?

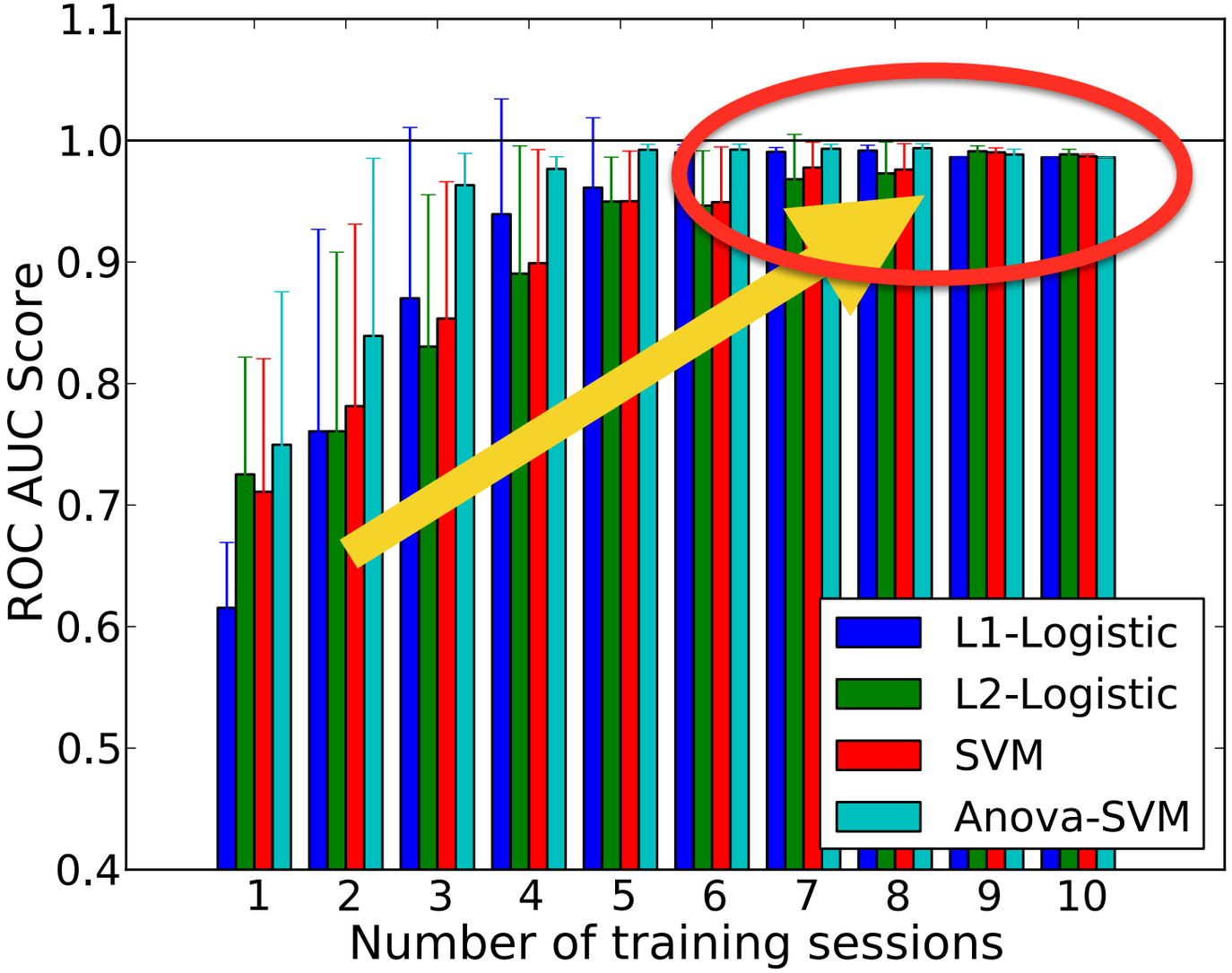


- 5 subjects
- 12 sessions (more than 1000 scans)
- Binary classification (face vs. house)
- Test of 2 left-out sessions

The more data the better

Data from *[Haxby et al. 2001]*
Figure from *[Gramfort et al. 2011]*

Why more data is better?



- 5 subjects
- 12 sessions (more than 1000 scans)
- Binary classification (face vs. house)
- Test of 2 left-out sessions

The more data the better

Almost 100% (no noise)

Data from *[Haxby et al. 2001]*
Figure from *[Gramfort et al. 2011]*

Problem:
"big data" in science
is generally
unsupervised

Project 1

- Objective: Learning representations from neural time series with self-supervision and data augmentation

Project 1

- Objective: Learning representations from neural time series with self-supervision and data augmentation

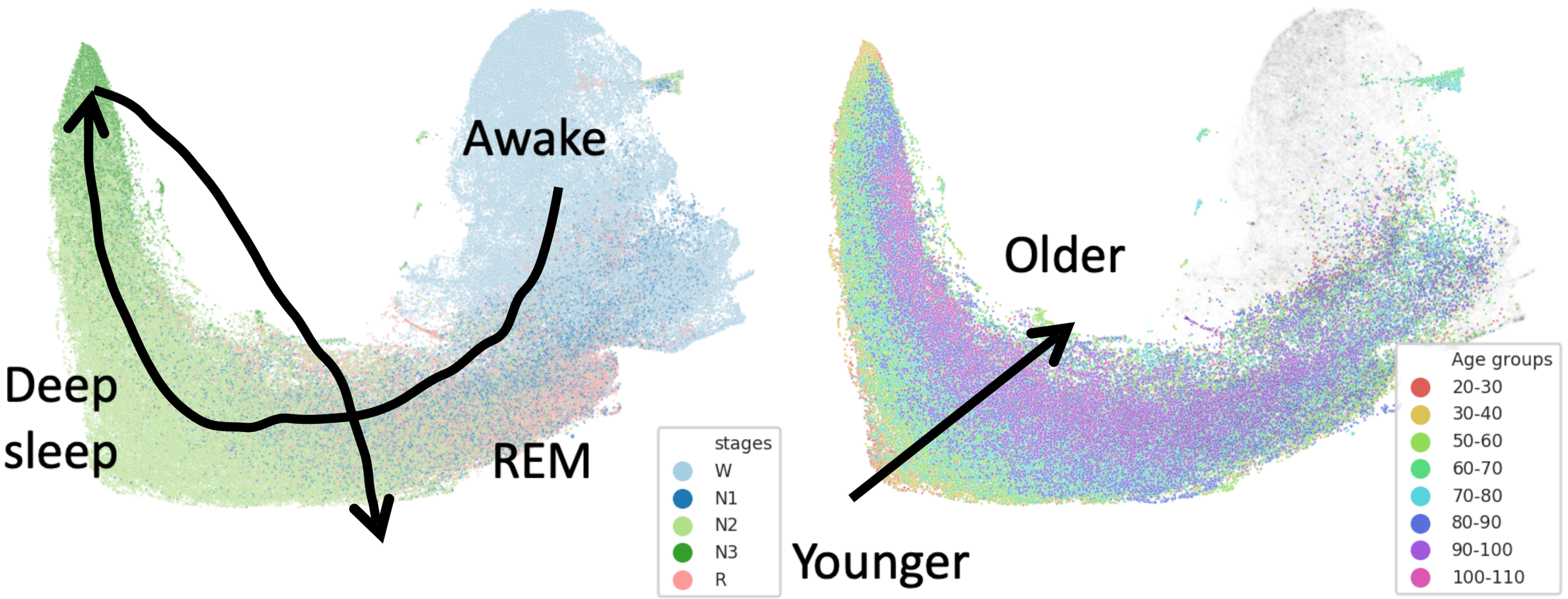
E.g.: **Jigsaw puzzle task** from Noroozi & Favaro (2016)

In a nutshell: use the **structure** of the data to pretrain a **feature extractor** with a supervised ("pretext") task – then use the features.

Other examples: word2vec, BERT, nonlinear ICA, etc.

Project 1

- Objective: Learning representations from neural time series with self-supervision and data augmentation



[Banville et al. MLSP 2019]

SSL to learn on sleep EEG

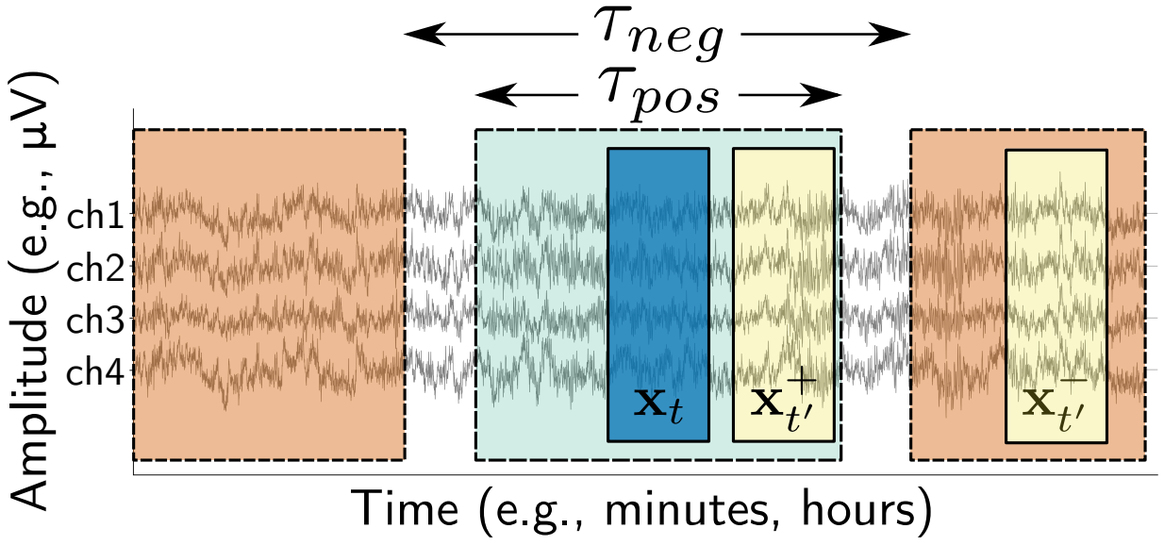
Problem:
What pretext task makes sense for EEG/MEG?

- Use knowledge about sleep (slow cycles)
- Theoretical approaches based on recent results on identifiability of non-linear ICA

Possible Self Sup. Tasks

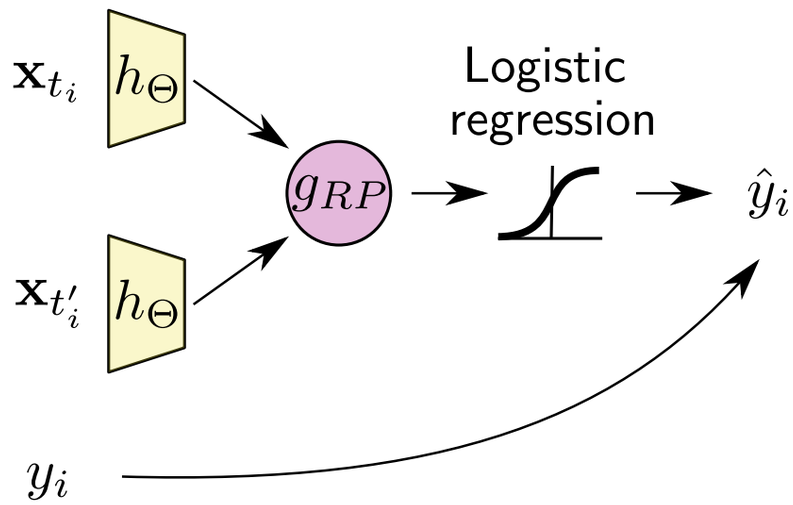
Relative positioning (RP)

1 Sampling



$$y_i = \begin{cases} 1, & \text{if } |t_i - t_i'| \leq \tau_{pos} \\ \text{-1}, & \text{if } |t_i - t_i'| > \tau_{neg} \end{cases}$$

2 Training

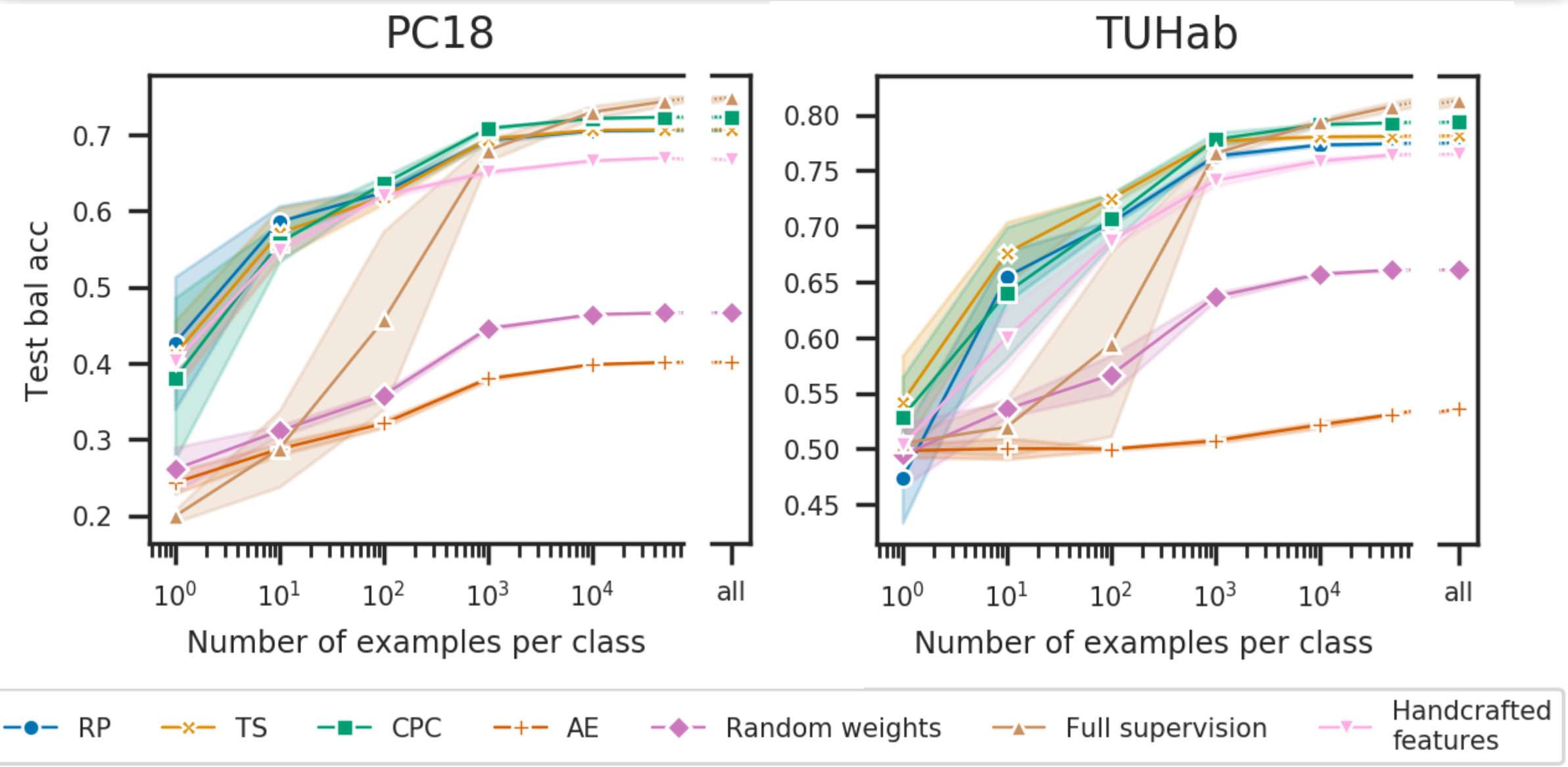


Predict if 2 windows of data are close in time

Other approaches: CPC [Oord et al. 2018], PCL [Hyvärinen et al. 2017] etc.

Project 1

- Objective: Learning representations from neural time series with self-supervision and data augmentation



Problem: Augmenting MEG/EEG data is not as simple as for images or speech

- Use the physics of MEG/EEG
- Use knowledge/availability of pure noise
- Use knowledge about neuroscience (freq. shifts, biophysiological models)

Problem:
Augmenting MEG/EEG data is not as simple as for images or speech

- Use the physics of MEG/EEG
- Use knowledge/availability of pure noise
- Use kno… …q.

shifts, bi…

We want to learn how to augment neuroscience data!

Problem of dataset variability

- ≠ recording devices / scanners
- ≠ EEG channels / fMRI sequence parameters
- ≠ preprocessing steps
- ≠ populations: ages, sexes, clinical disorders…
- ≠ labeling guidelines

Problem of dataset variability

- ≠ recording devices / scanners
- ≠ EEG channels / fMRI sequence parameters
- ≠ preprocessing steps
- ≠ populations: ages, sexes, clinical disorders…
- ≠ labeling guidelines

- Pooling datasets to increase n can reduce performance
- Performance on new dataset can drop

Problem of dataset variability

- ≠ recording devices / scanners
- ≠ EEG channels / fMRI sequence parameters
- ≠ preprocessing steps
- ≠ populations:
- ≠ labeling guide

- Pooling datasets to
- Performance on n

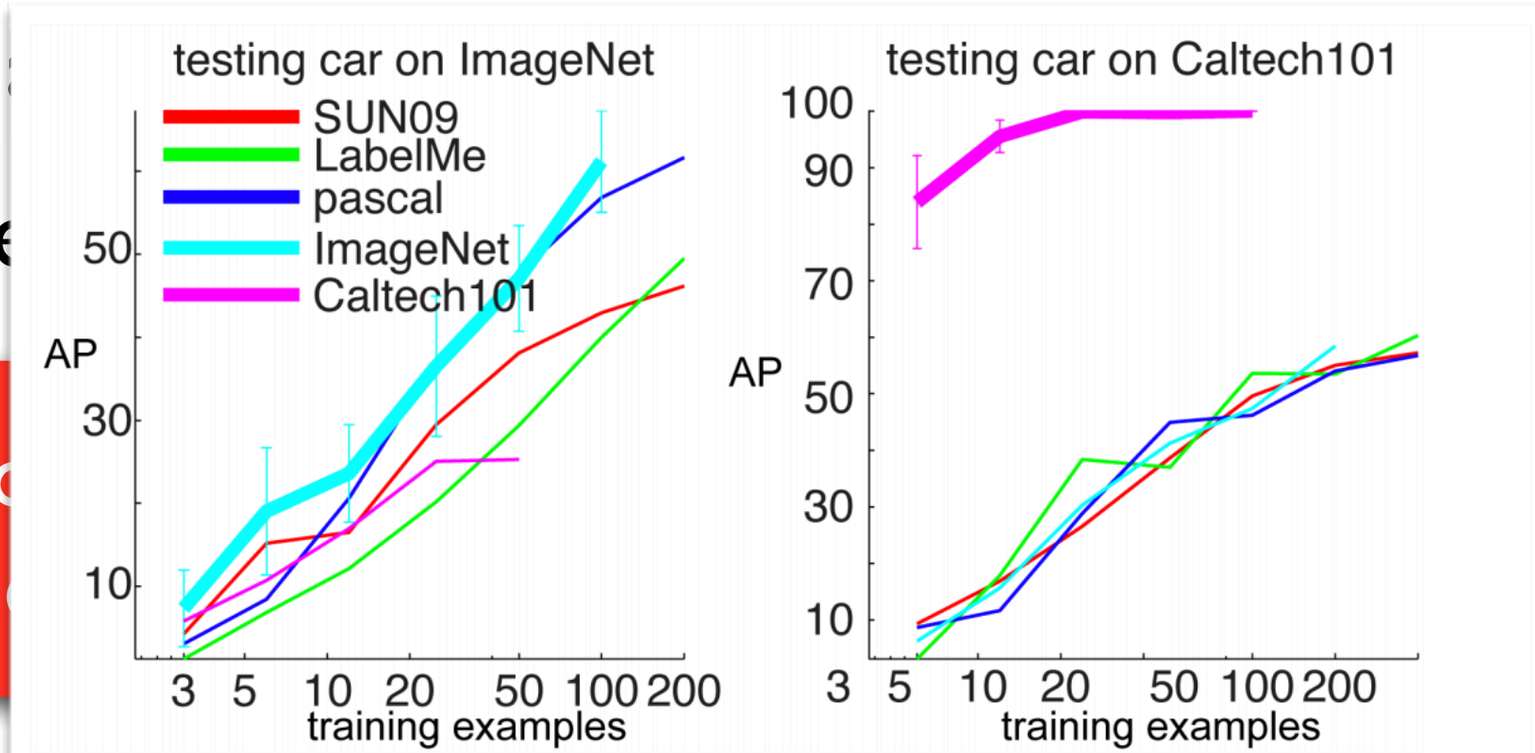
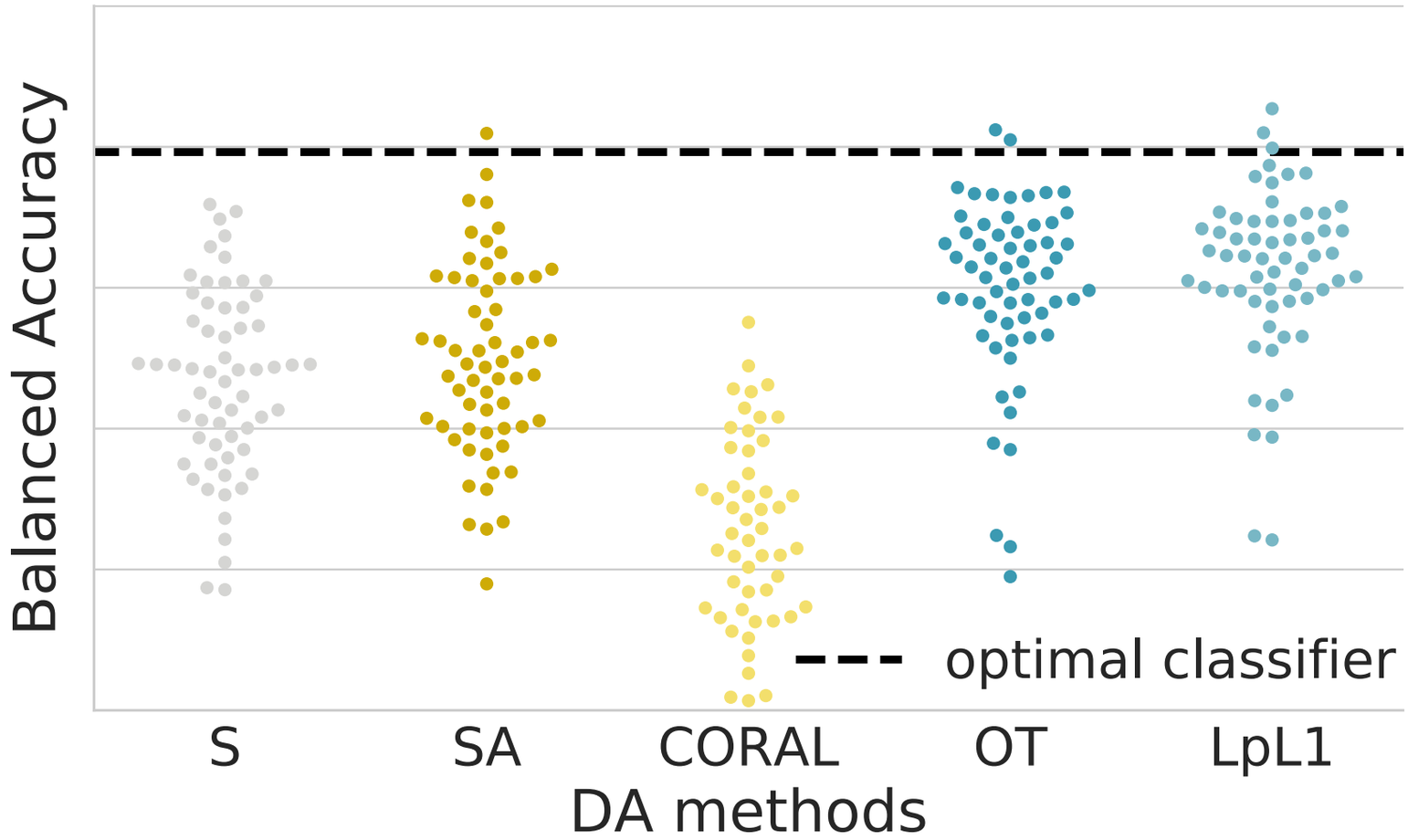


Figure 6. Cross-dataset generalization for "car" detection as function of training data

[Torralba and Efros, 2011]

Domain adaptation with EEG sleep

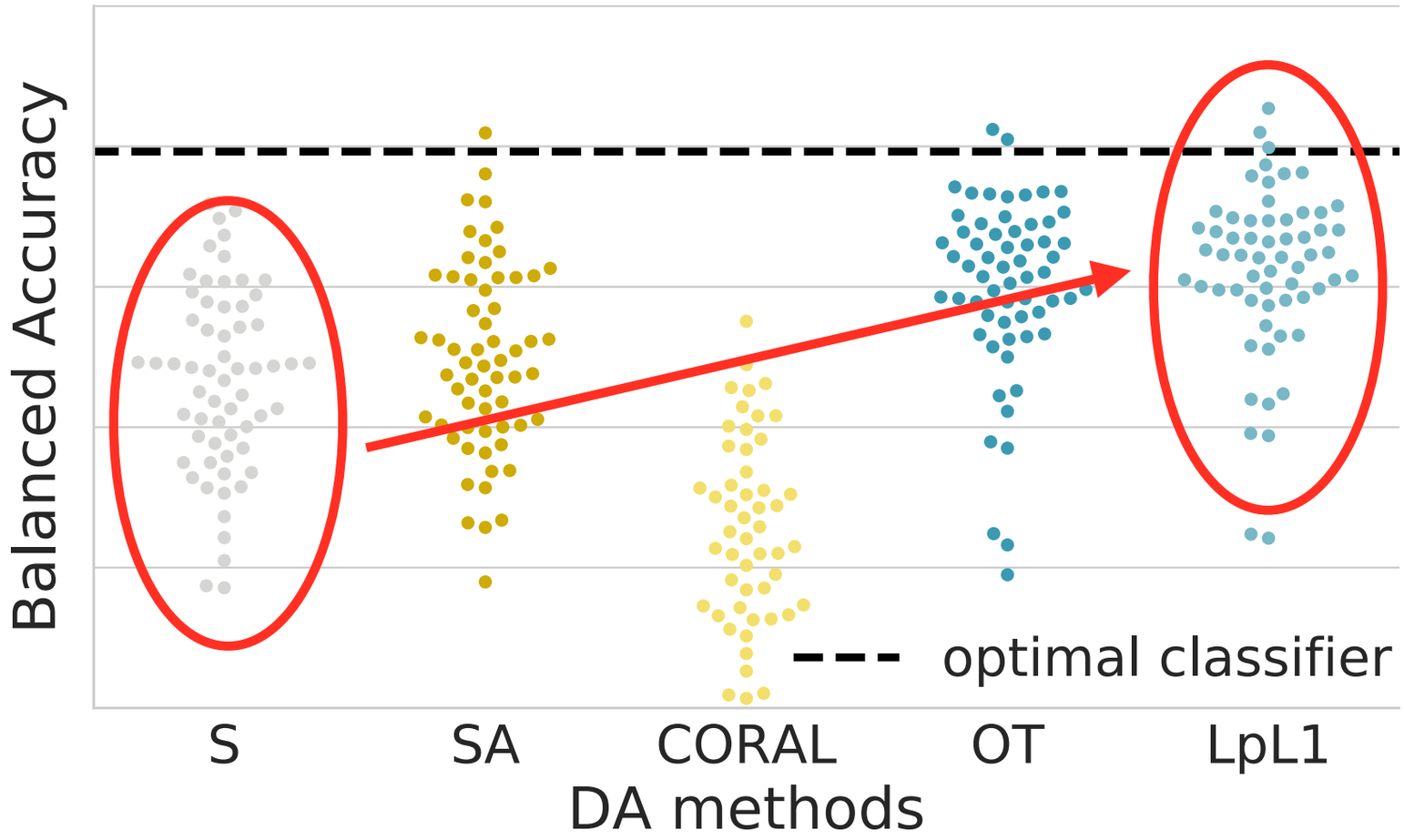
- Train dataset: MESA *[Dean et al. 2016]*
- Test dataset: MASS-session 3 *[O'Reilly et al. 2014]*
- 3 EEG + 2 EOG channels



[Chambon et al., Domain adaptation with optimal transport improves EEG sleep stage classifiers, PRNI 2018]

Domain adaptation with EEG sleep

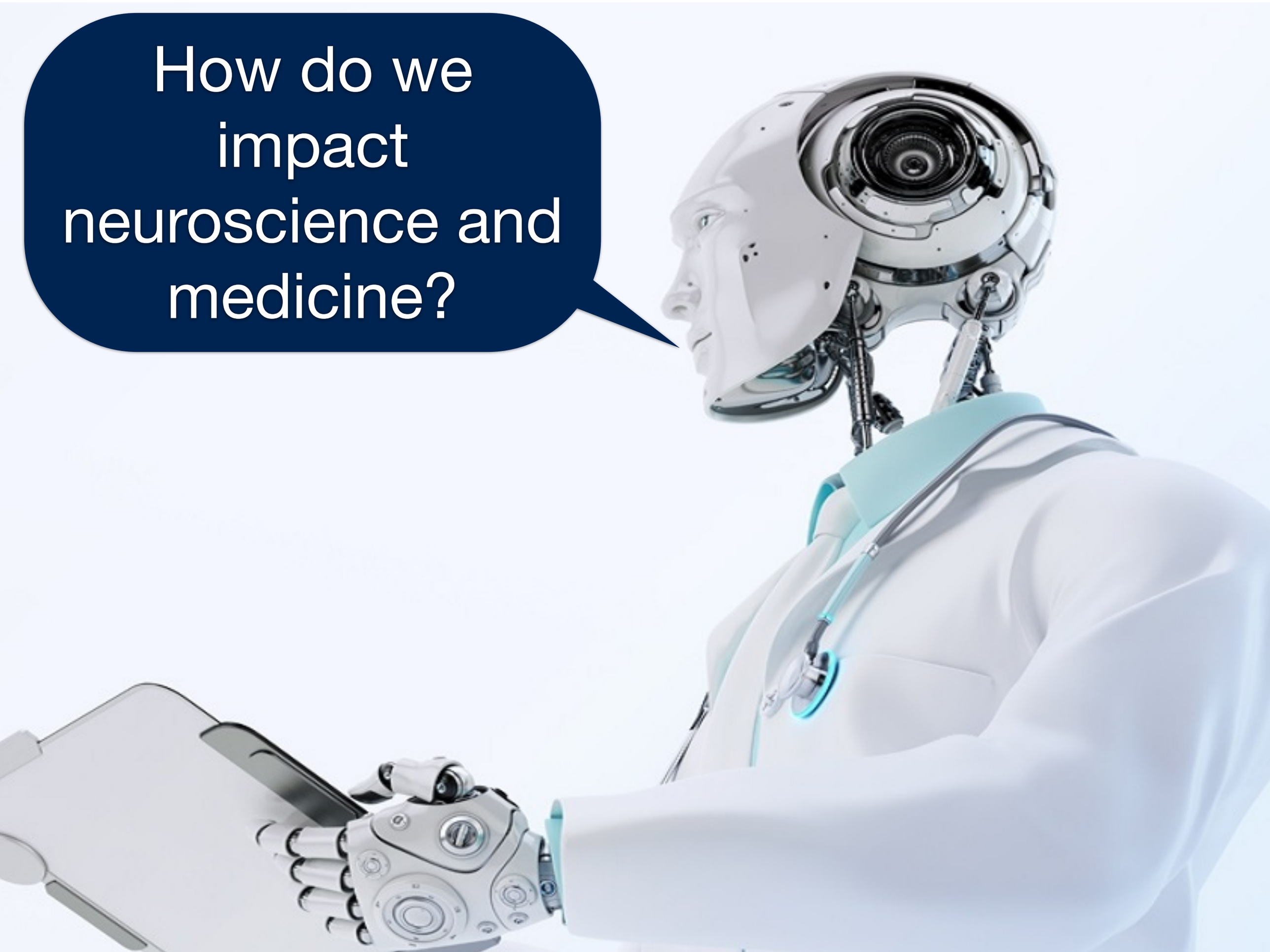
- Train dataset: MESA *[Dean et al. 2016]*
- Test dataset: MASS-session 3 *[O'Reilly et al. 2014]*
- 3 EEG + 2 EOG channels



Domain adaptation improves performance

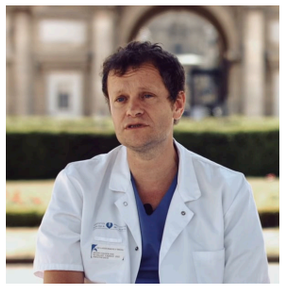
[Chambon et al., Domain adaptation with optimal transport improves EEG sleep stage classifiers, PRNI 2018]

How do we impact neuroscience and medicine?



Predict of brain “fragility” for optimal drug dosage across age

Joint work with:



informatiques mathématiques

Inria

Inserm

Institut national
de la santé et de la recherche médicale

Hôpitaux Universitaires

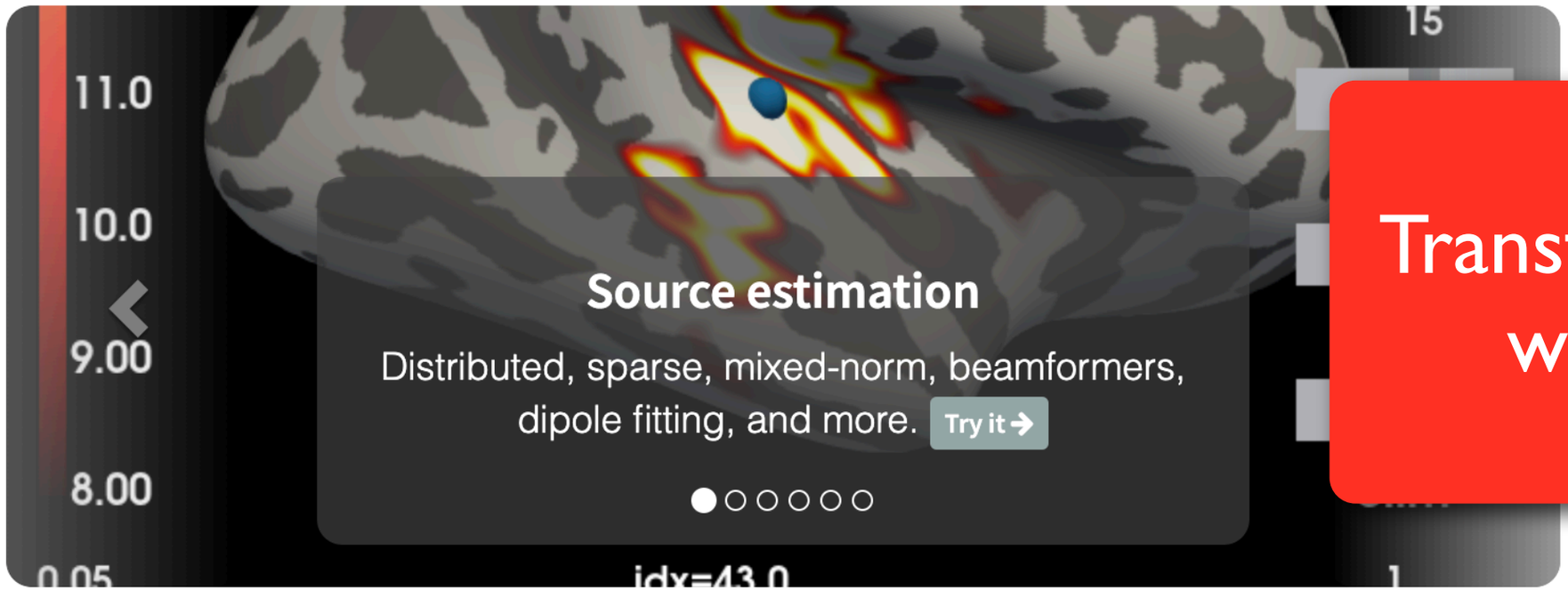
GH

LARIBOISIÈRE

MEG + EEG ANALYSIS & VISUALIZATION

Open- analyzing
huma RS, and

https://mne.tools/



Transfer + impact with MNE

Version 0.21.dev0

- What's new
- Installation
- Documentation
- Cite

Direct financial support

National Institutes of Health: R01-EB009048, EB009048, EB006385, HD40712, NS44319, NS37462, NS104585, **P41**-EB015896, RR14075-06

US National Science Foundation: 0958669, 1042134

European Research Council: YStG-263584, 676943

E-FG02-

erche:

03-02

Science:

s

Chan Zuckerberg Initiative: Essential Open Source Software for Science

MNE software for processing MEG and EEG data, A. Gramfort, M. Luessi, E. Larson, D. Engemann, D. Strohmeier, C. Brodbeck, L. Parkkonen, M. Hämäläinen, Neuroimage 2013

Objectives

BrAIN objective: Develop the next ML paradigms to extract knowledge from physiological signals

O1. Learn with no-supervision on noisy and complex multivariate signals

O2. Learn end-to-end predictive systems from limited data exploiting physical constraints

O3. Learn from data coming from many different source domains

O4. Develop high-quality software tools that can reach clinical research

Team

- Denis Engemann
- Thomas Moreau
- 1 Post-doc
- 1 Engineer
- 3 PhDs
- INSERM team at Larib. for clinical cases

- Aapo Hyvärinen as external collaborator/visitor

Contact

http://alexandre.gramfort.net

GitHub : @agramfort *Twitter : @agramfort*

informatics mathematics

"An approximate answer to the right problem is worth a good deal more than an exact answer to an approximate problem. ~ John Tukey"