Large-scale machine learning for medical imaging

Bertrand Thirion

bertrand.thirion@inria.fr

Outline

- Is machine learning useful for medical imaging ?
- Medical imaging in the big data era

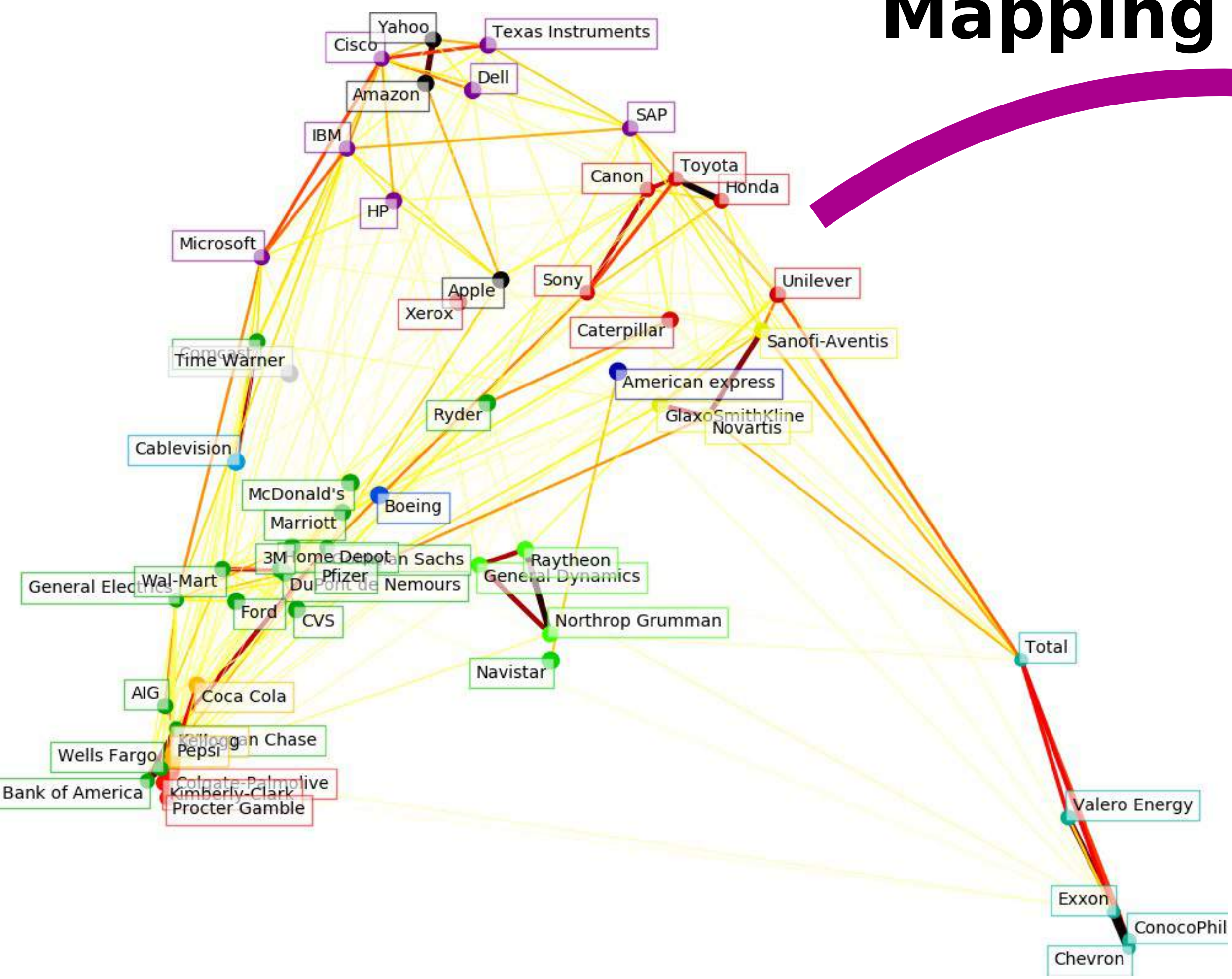
Outline

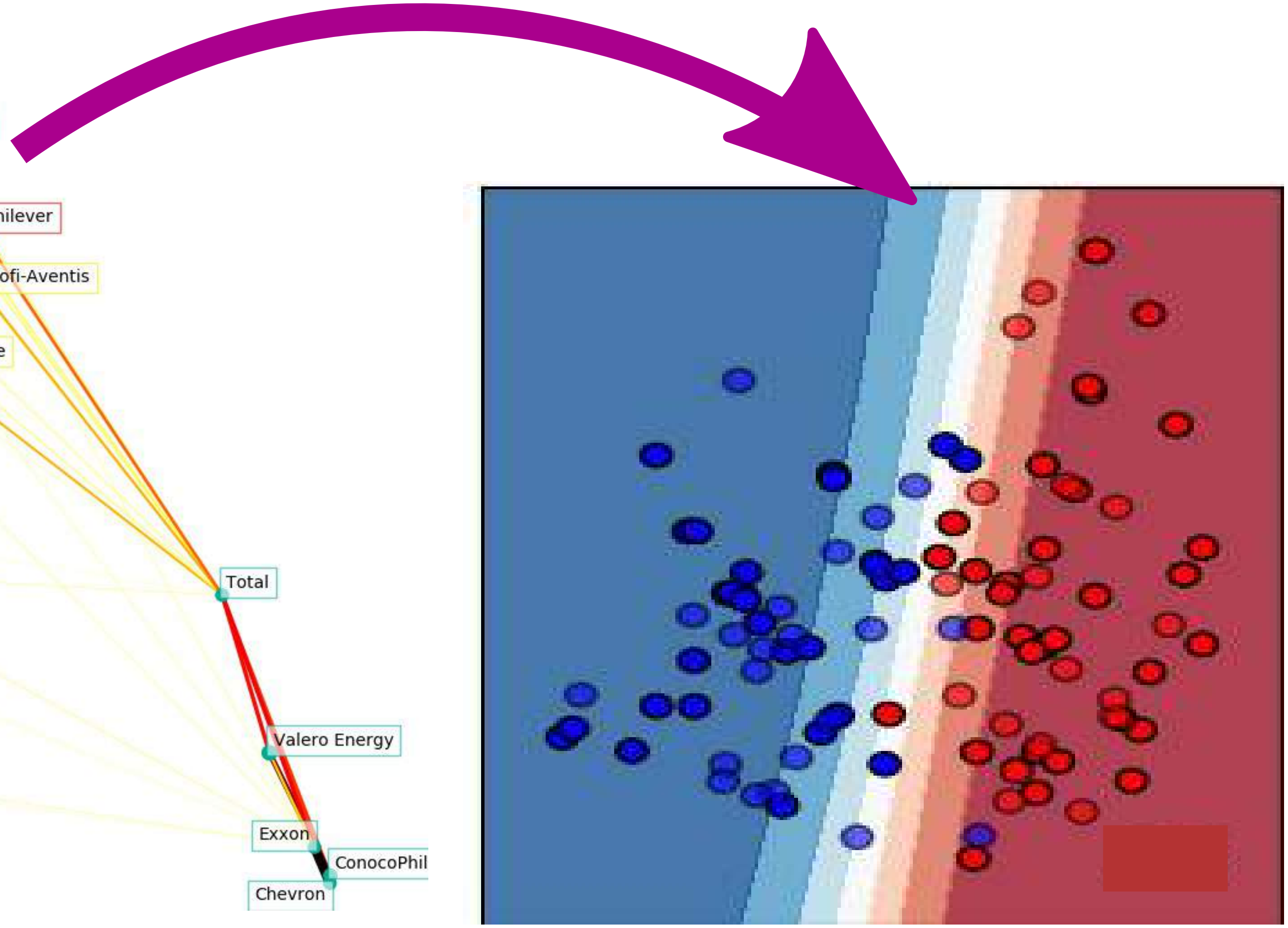
- Is machine learning useful for medical imaging ?
- Medical imaging in the big data era

What is machine learning good at ?

Supervised learning

Mapping F(x)

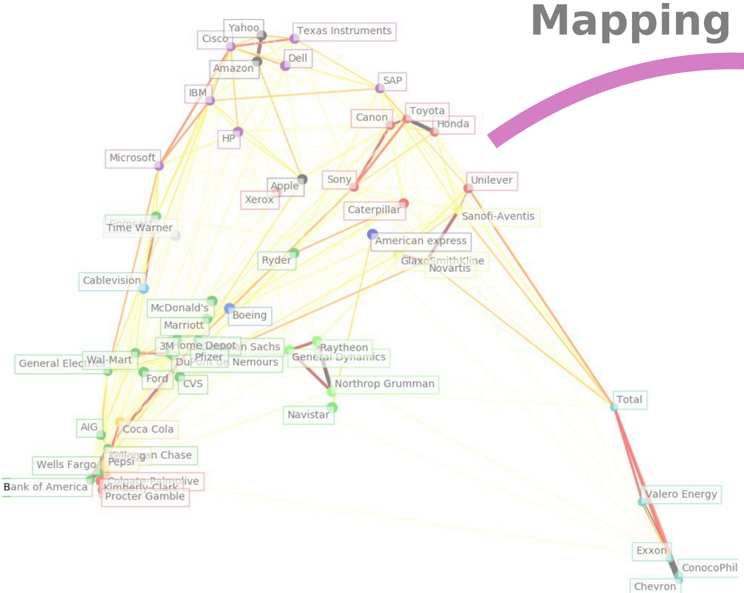




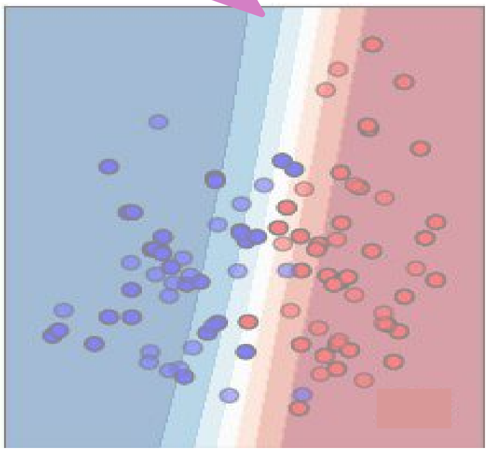
Space X (input objects) **Space Y (Labels)**

What is machine learning good at ?

Supervised learning



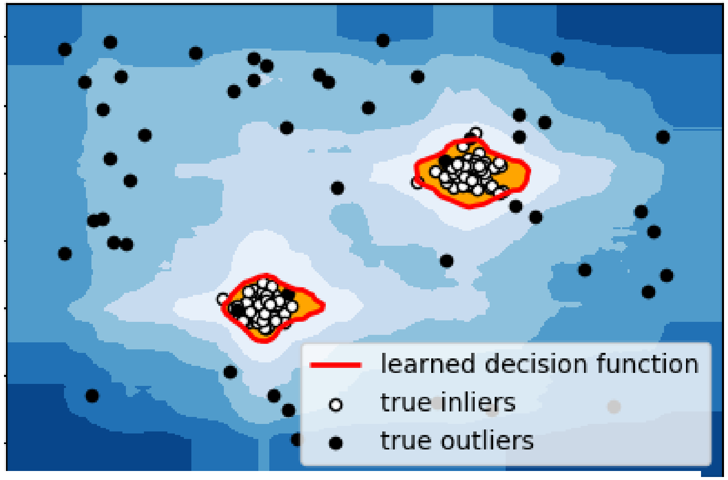
Mapping F(x)

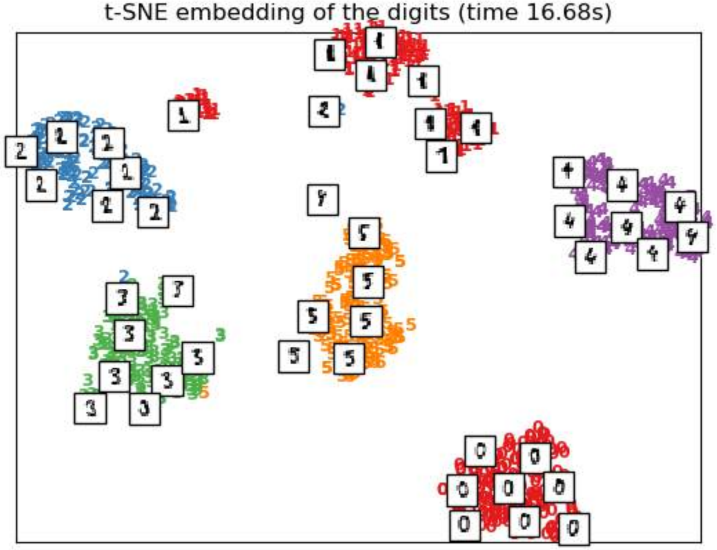


Space X (input objects)

Space Y (Labels)

Unsupervised learning

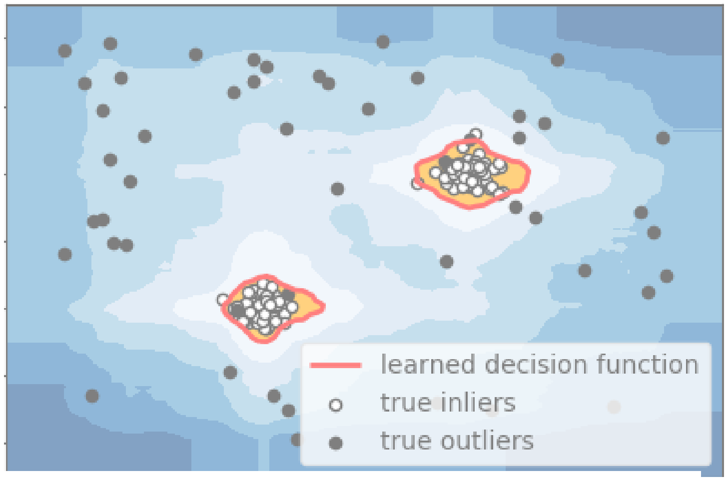




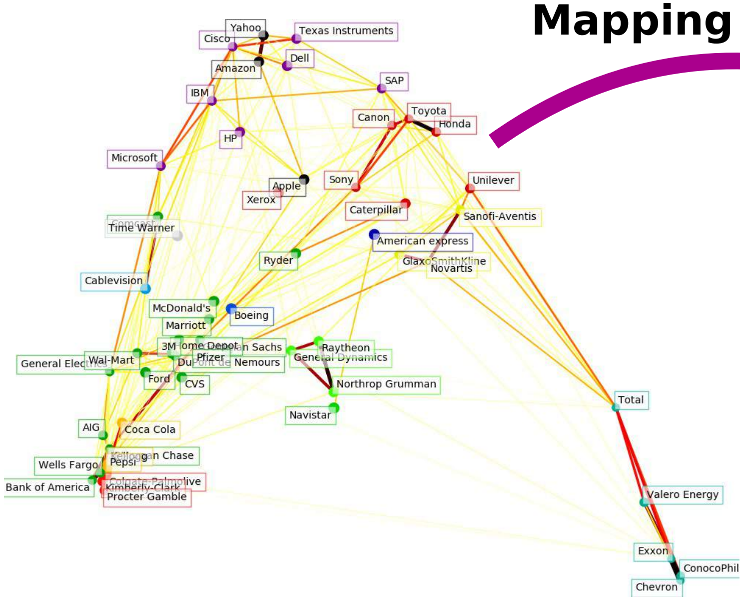
What is machine learning good at ?

✔ Supervised learning

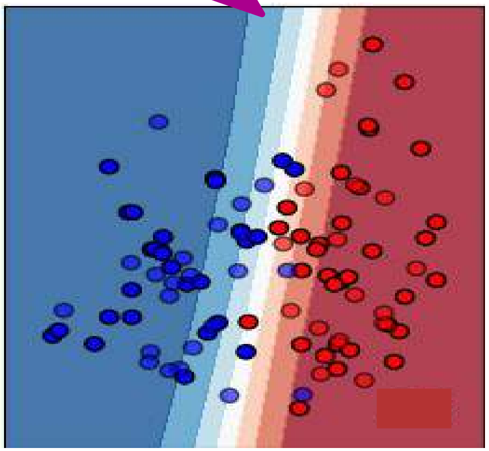
✗ Unsupervised learning



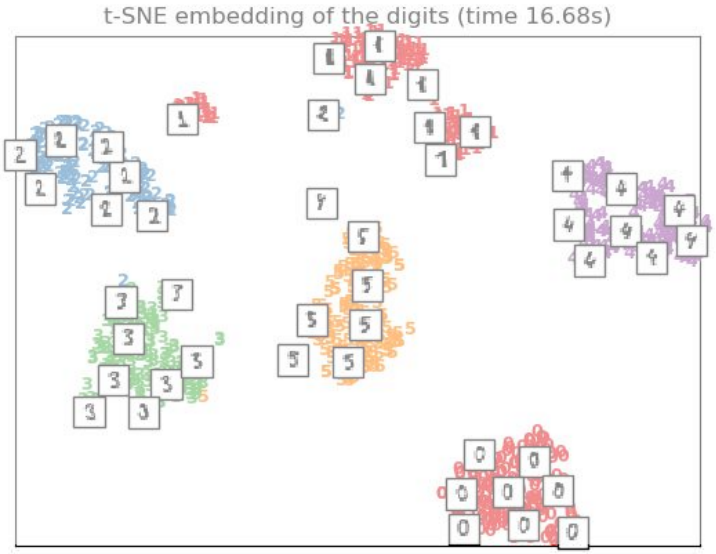
Mapping F(x)



Space X (input objects)

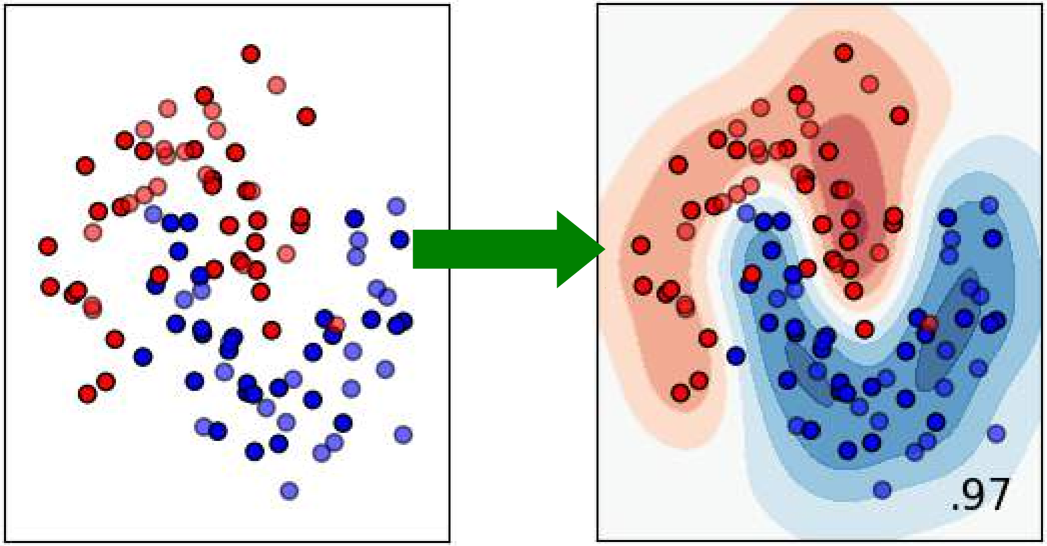


Space Y (Labels)



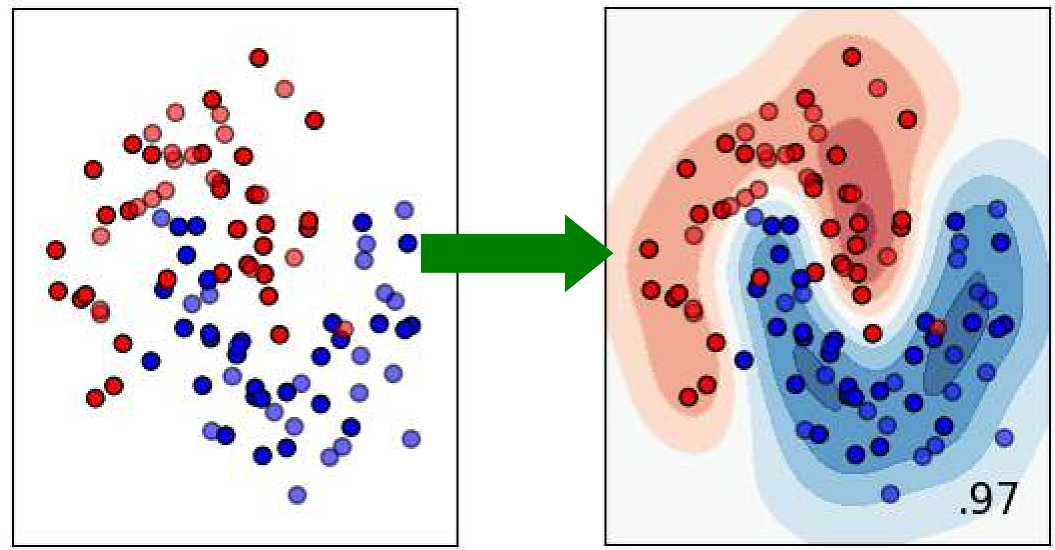
What is machine learning good at ?

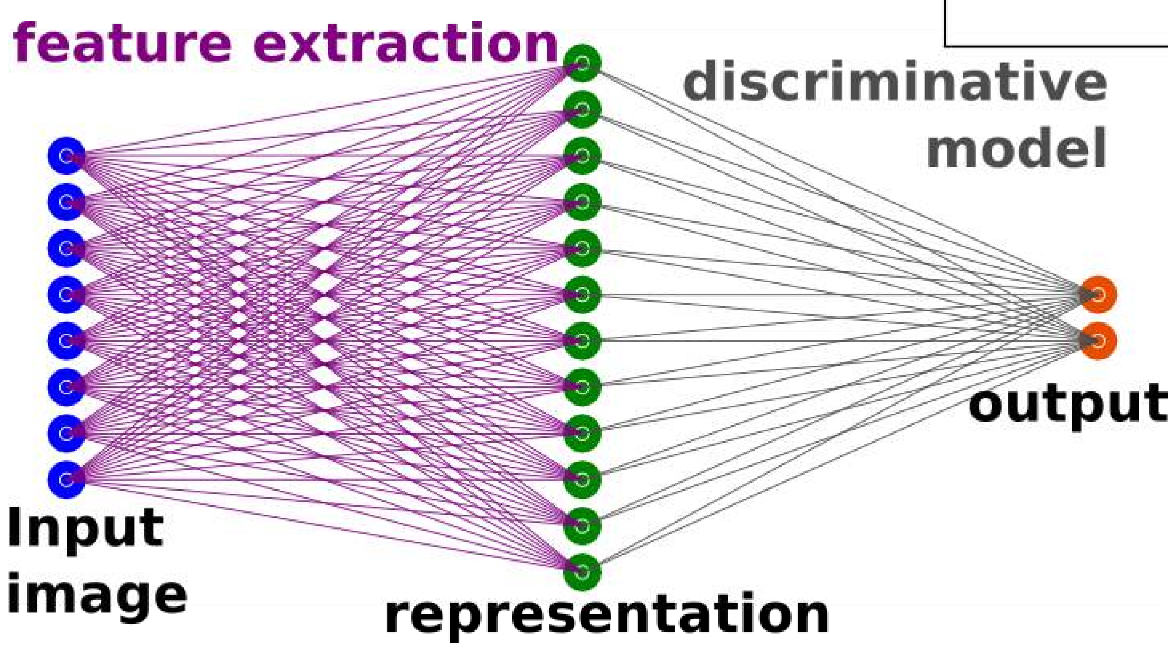
Supervised learning



What is machine learning good at ?

Supervised learning

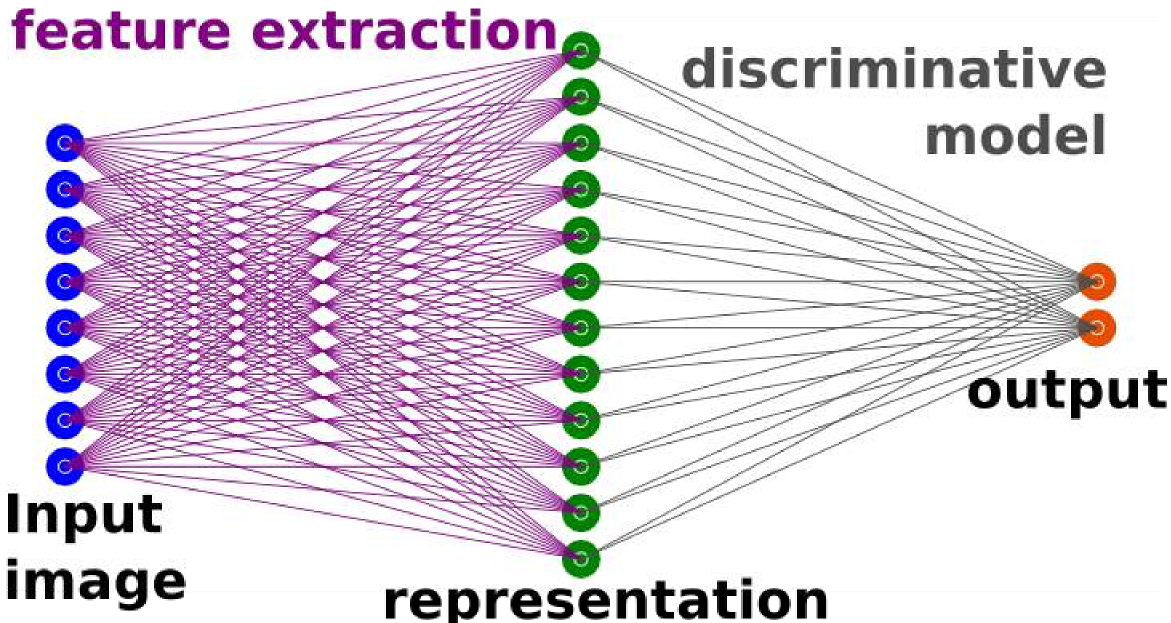




Feature extraction
= representation design
↔ kernel computation

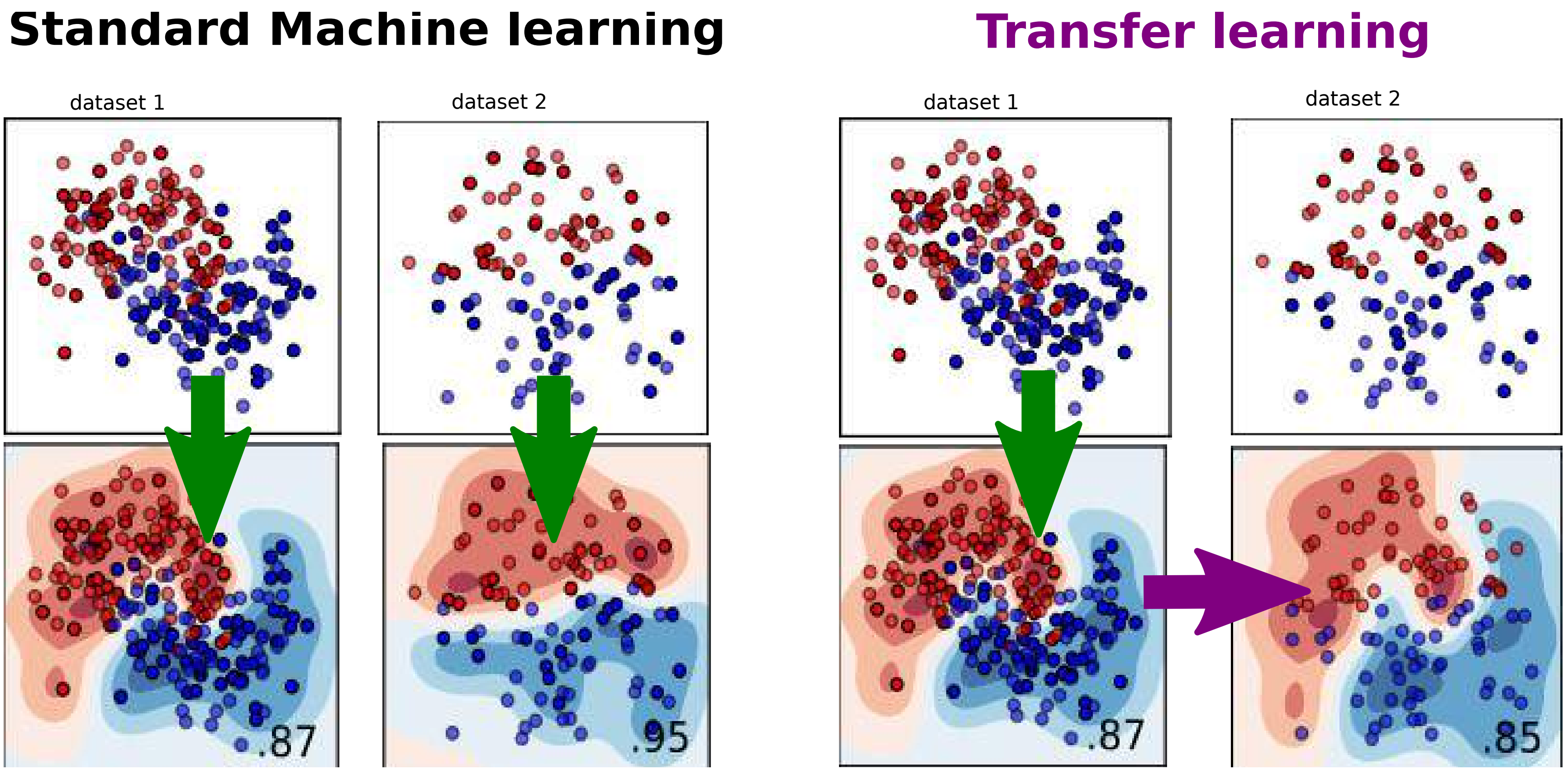
End-to-end learning

Done manually till recently *“modelling”:* mathematical morphology, wavelets, filtering, statistics



Currently done by the CNN alone
(possibly with transfer learning)

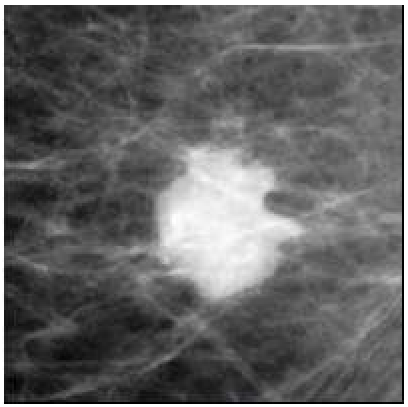
What is Machine Learning good at ?



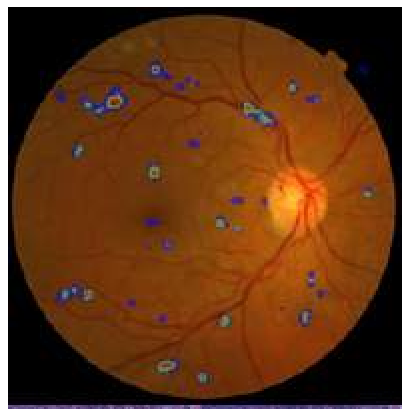
Transfer learning

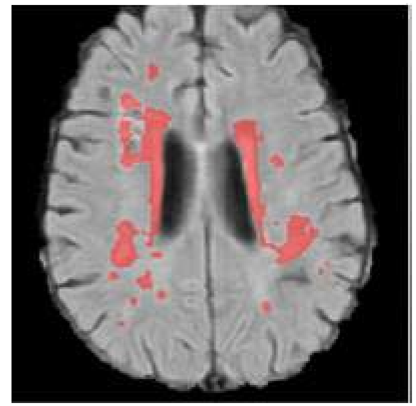
- stronger type of generalization
- successful transfer interesting as such

What medical imaging needs

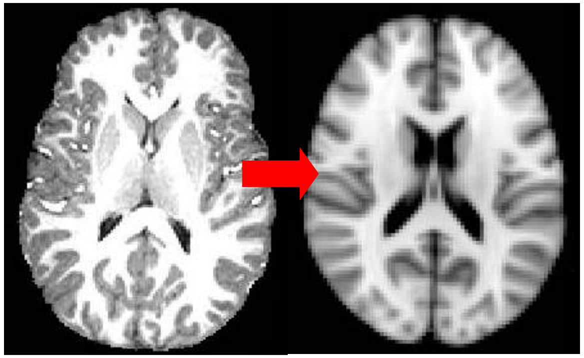


Classification





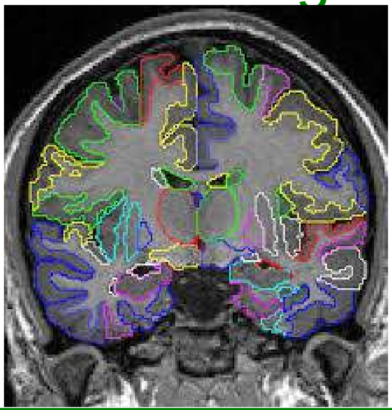
Detection

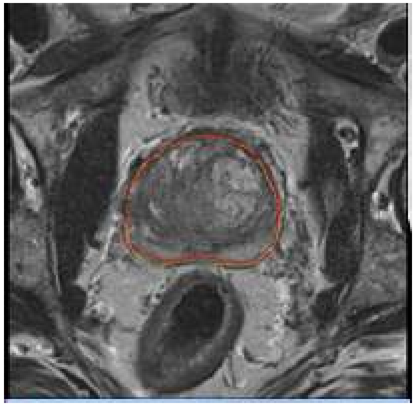


Registration

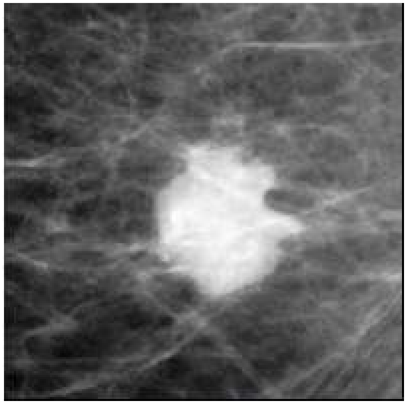
Image retrieval

Segmentation

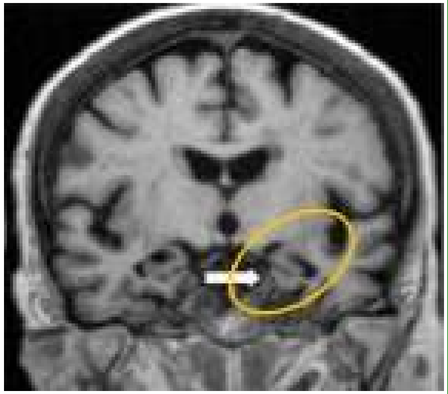


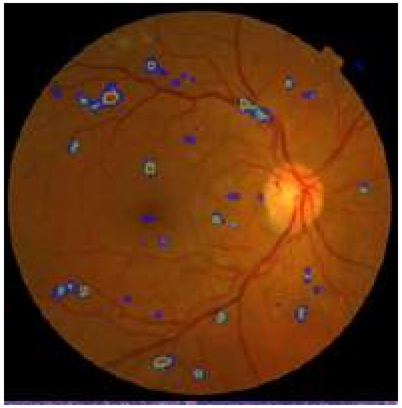


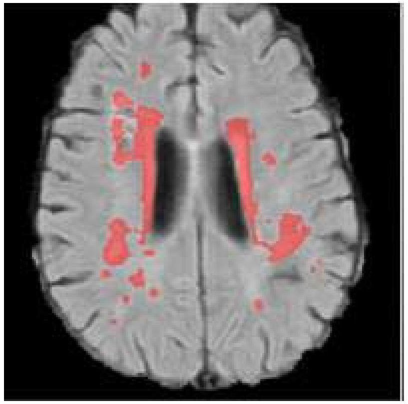
What medical imaging needs



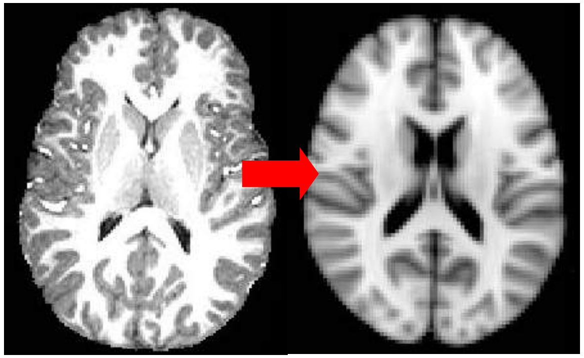
Classification







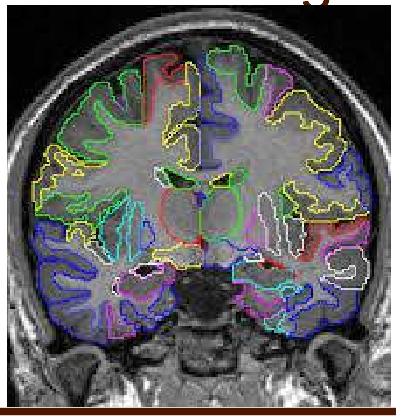
Detection

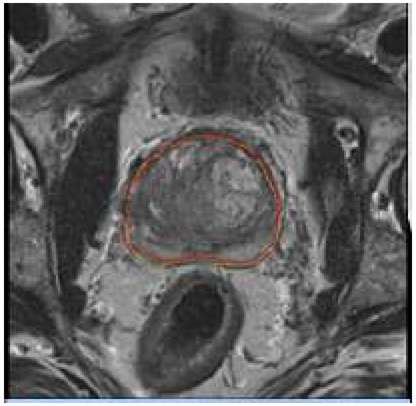


Registration

Image retrieval

Segmentation





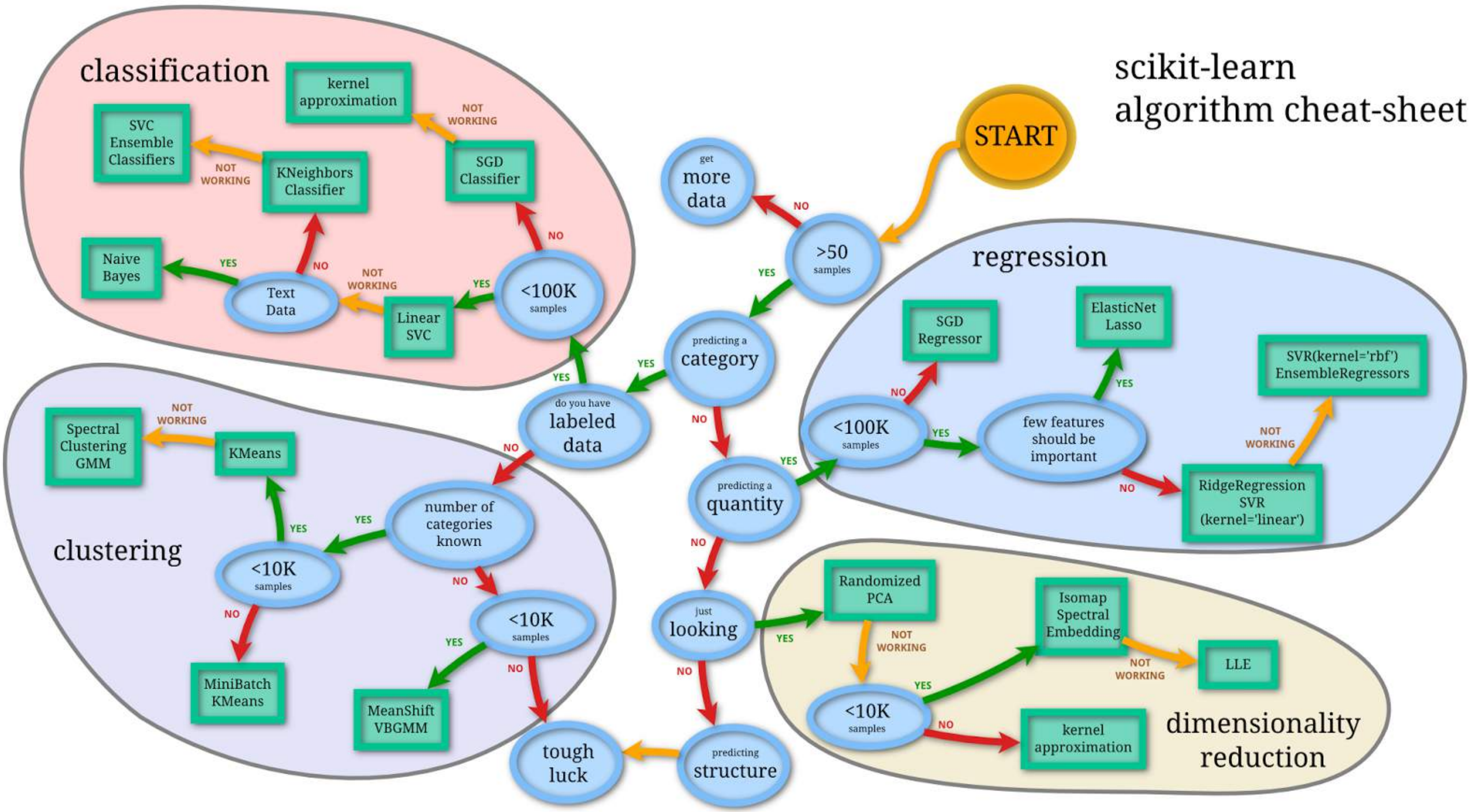
Outline

- Is machine learning useful for medical imaging ?
- Medical imaging in the big data era

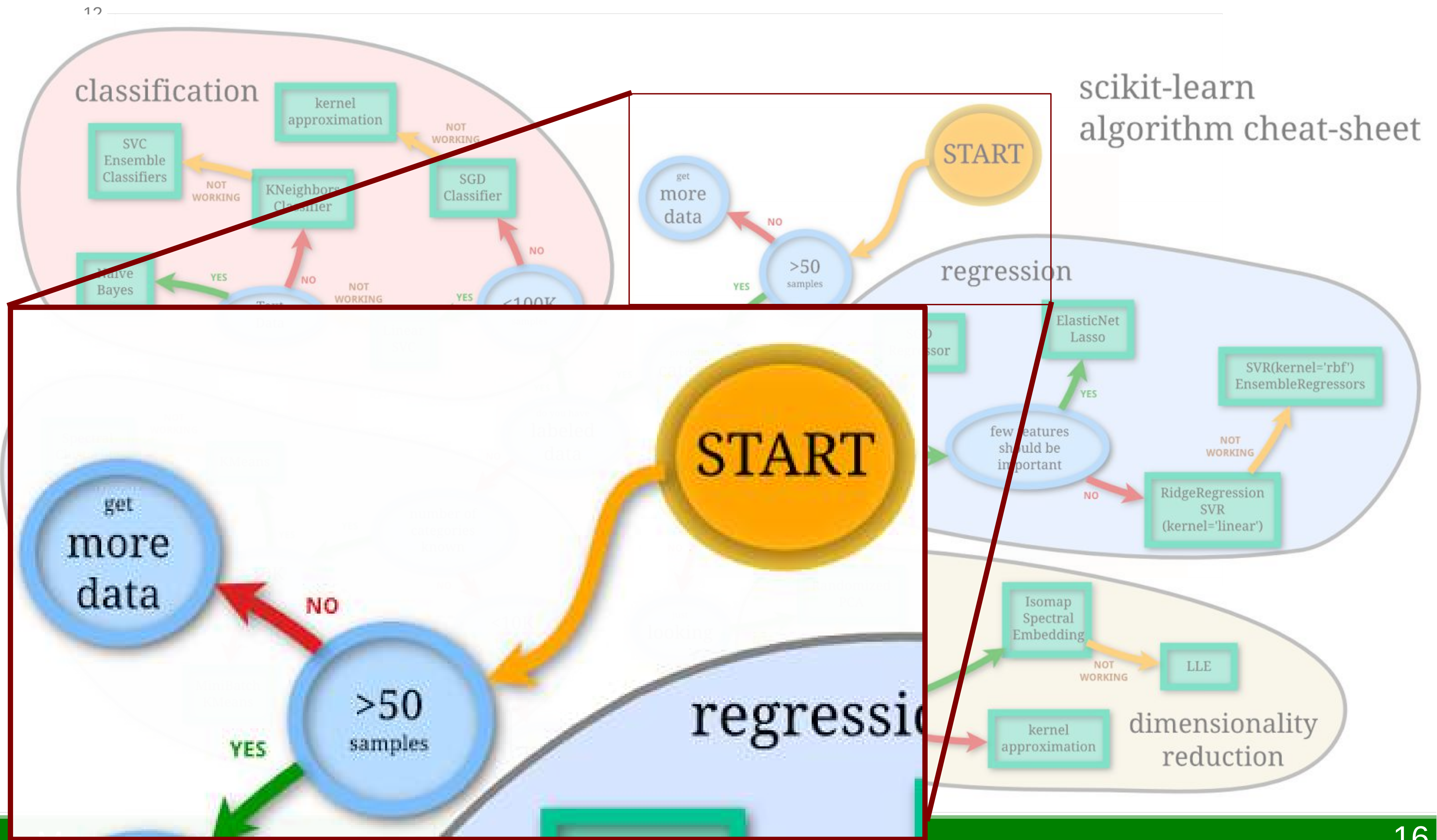
Medical Imaging in the big data era

- As any statistical analysis procedure, machine learning requires large sample sizes

Sample size & multivariate analysis

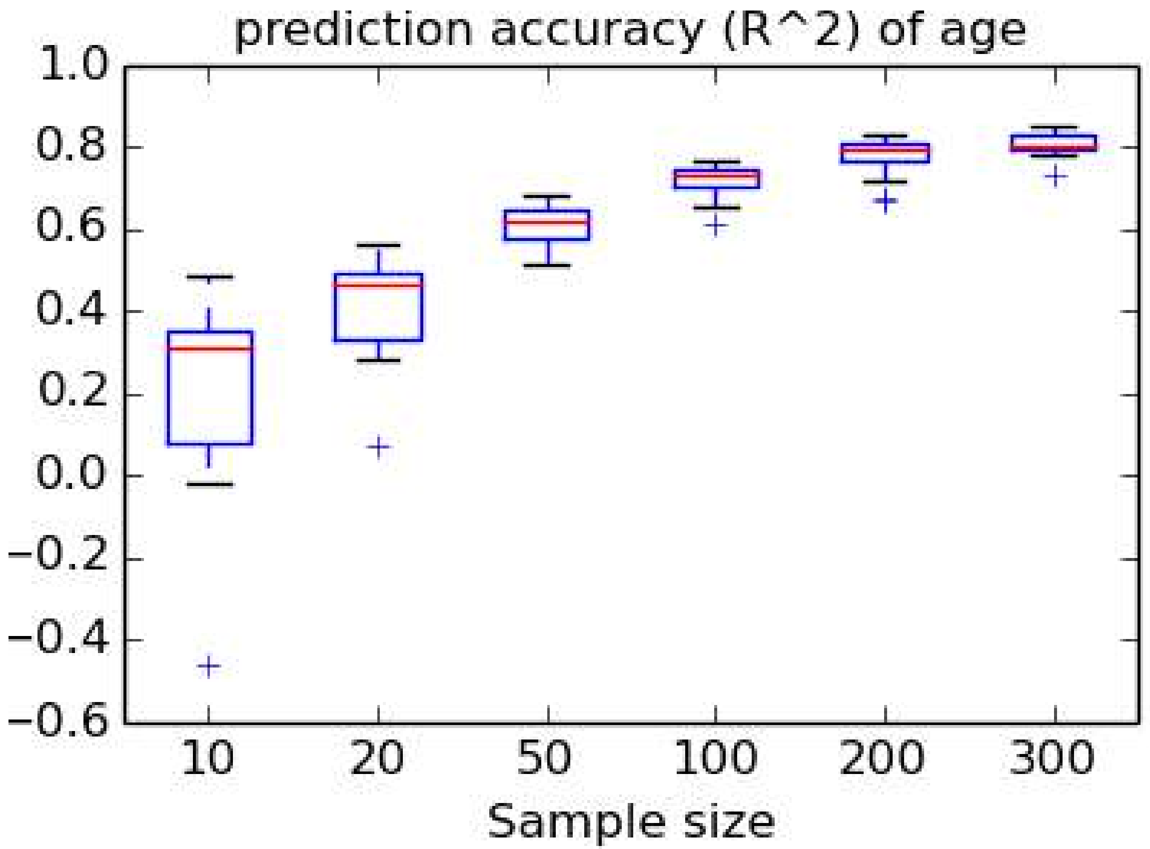


Multivariate analysis



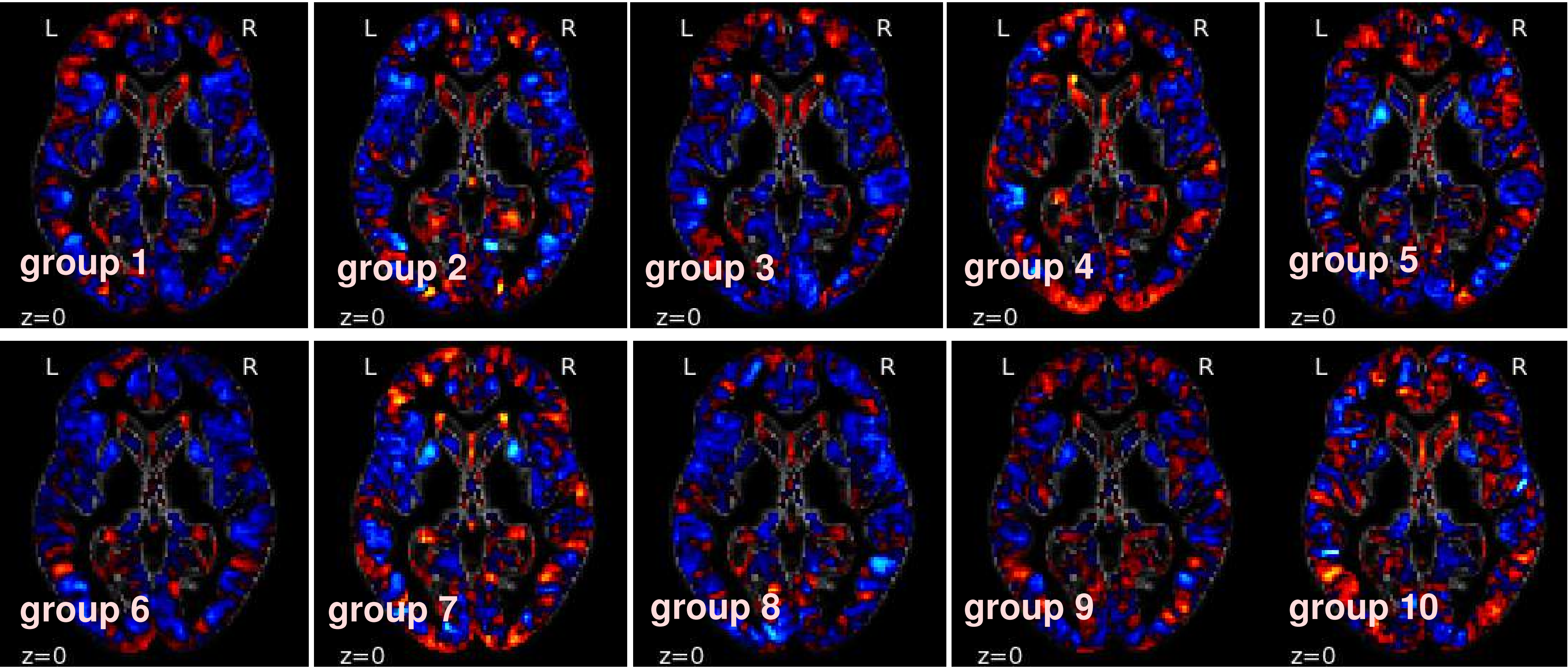
Learning curve: how prediction improves with n

- Predict the age of a subject given gray matter density maps (OASIS dataset, n=403)



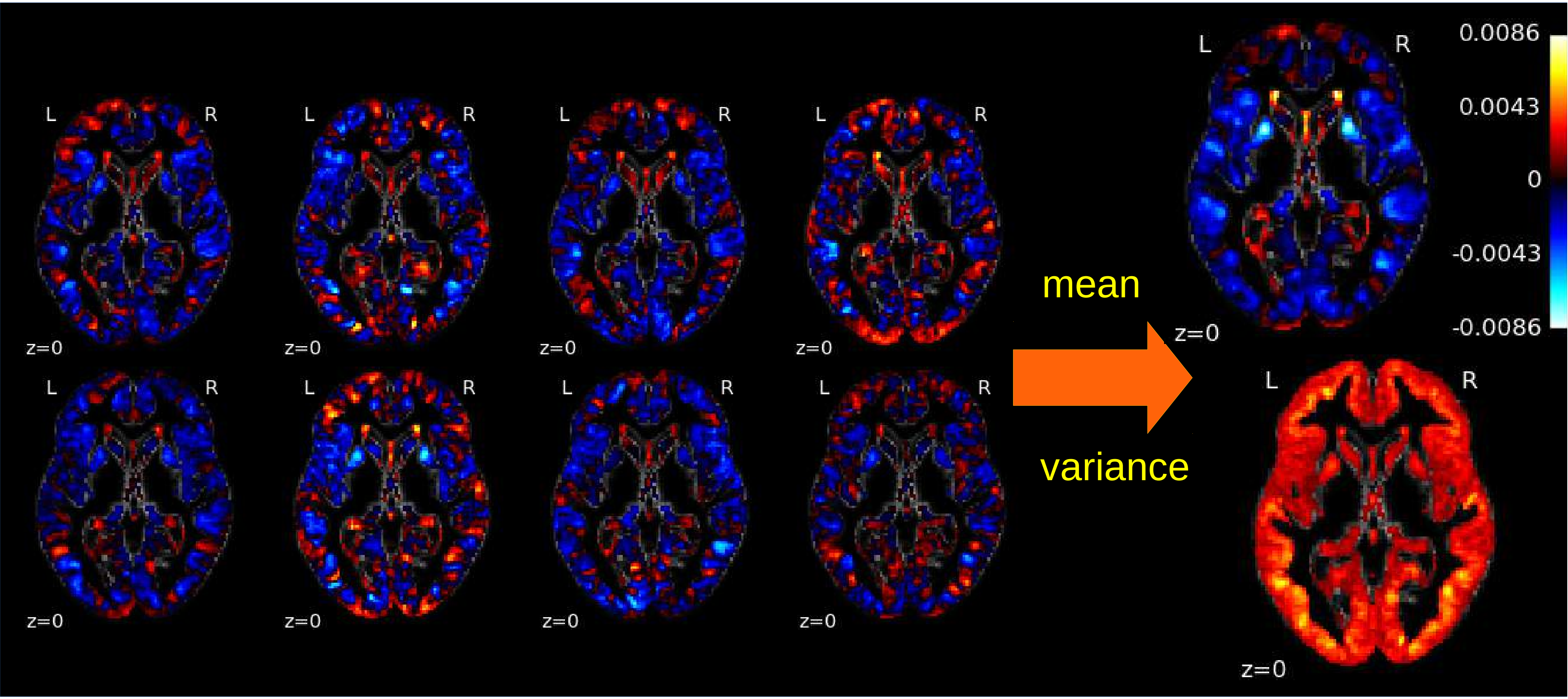
Weight maps for age prediction / OASIS

The weight map depends on the batch of subject considered (bootstrap):
One question, different datasets, different answers



Weight maps for age prediction / OASIS

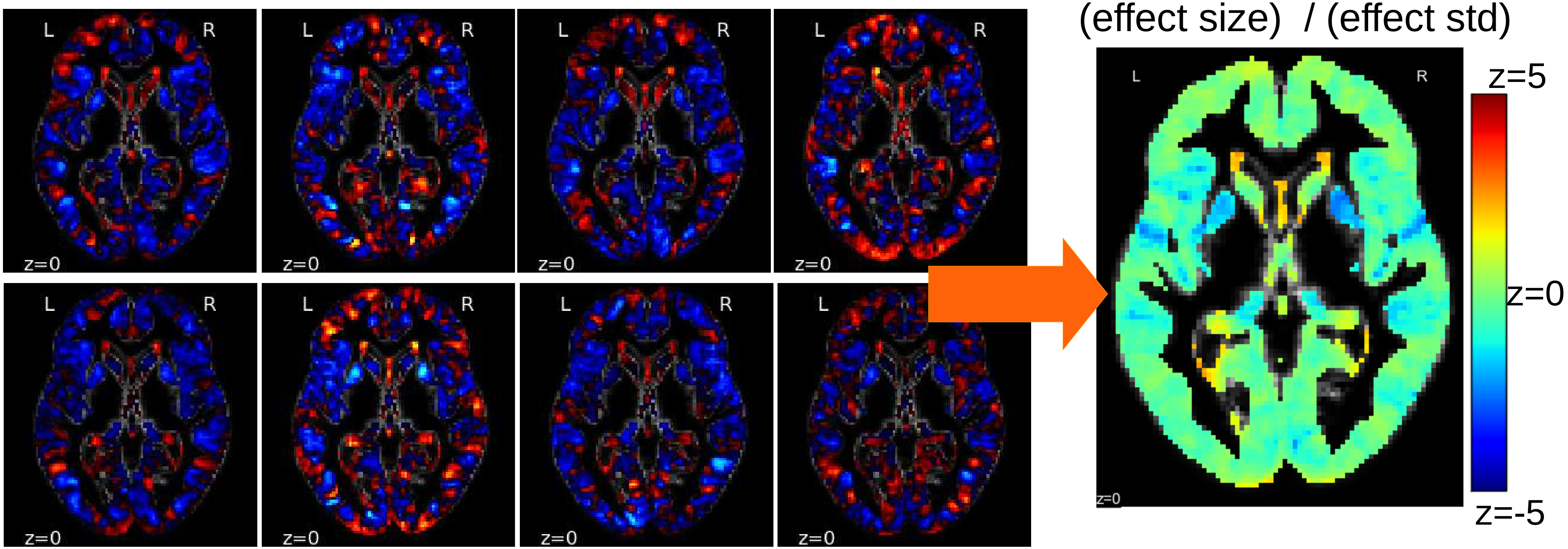
The weight map depends on the batch of subject considered (bootstrap):
One question, different datasets, different answers



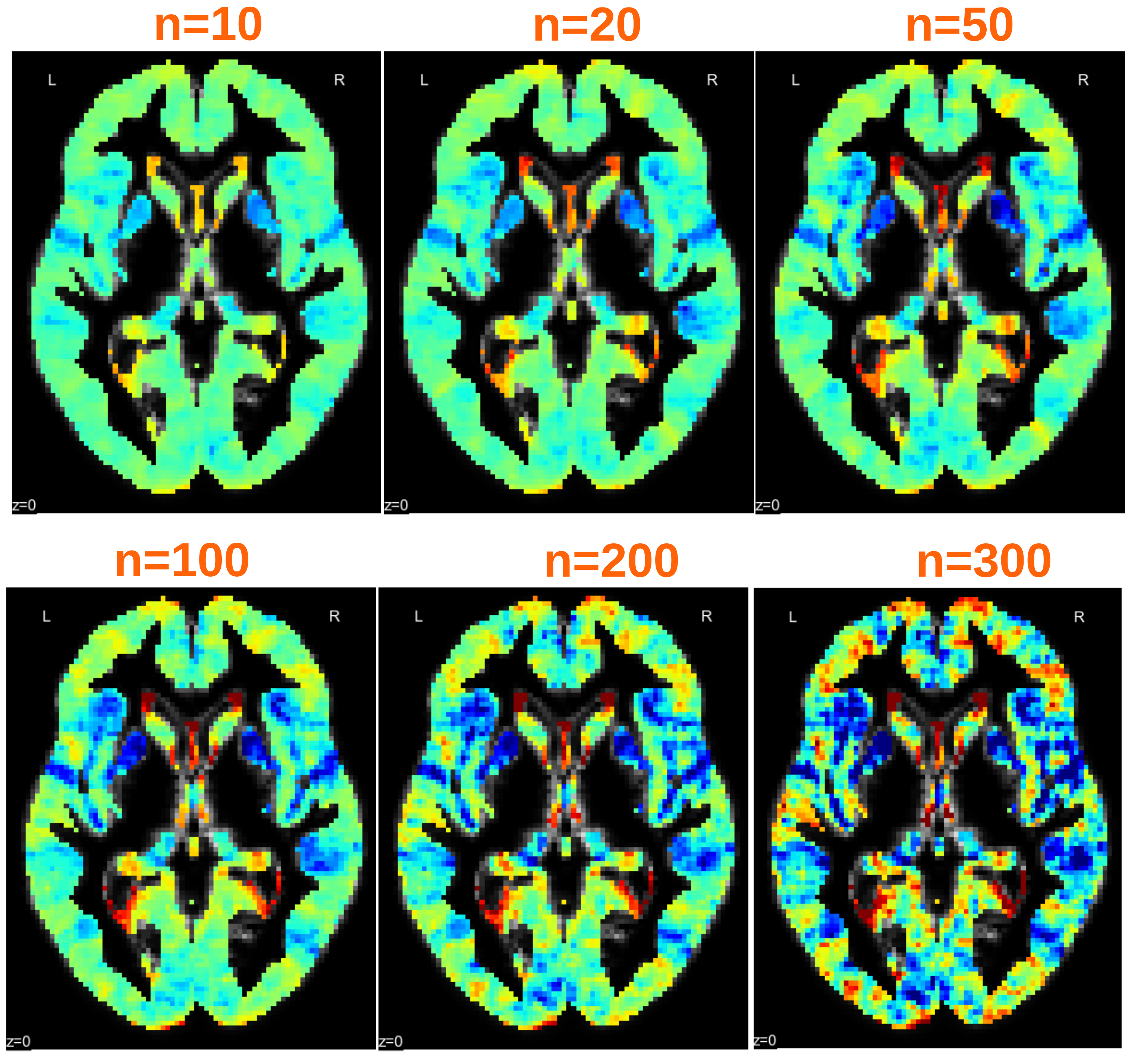
Weight maps for age prediction / OASIS

The weight map depends on the batch of subject considered (bootstrap):
One question, different datasets, different answers

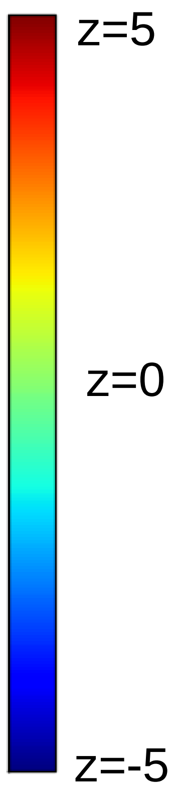
Summarized into a z image:
(effect size) / (effect std)



Weight maps for age prediction / OASIS

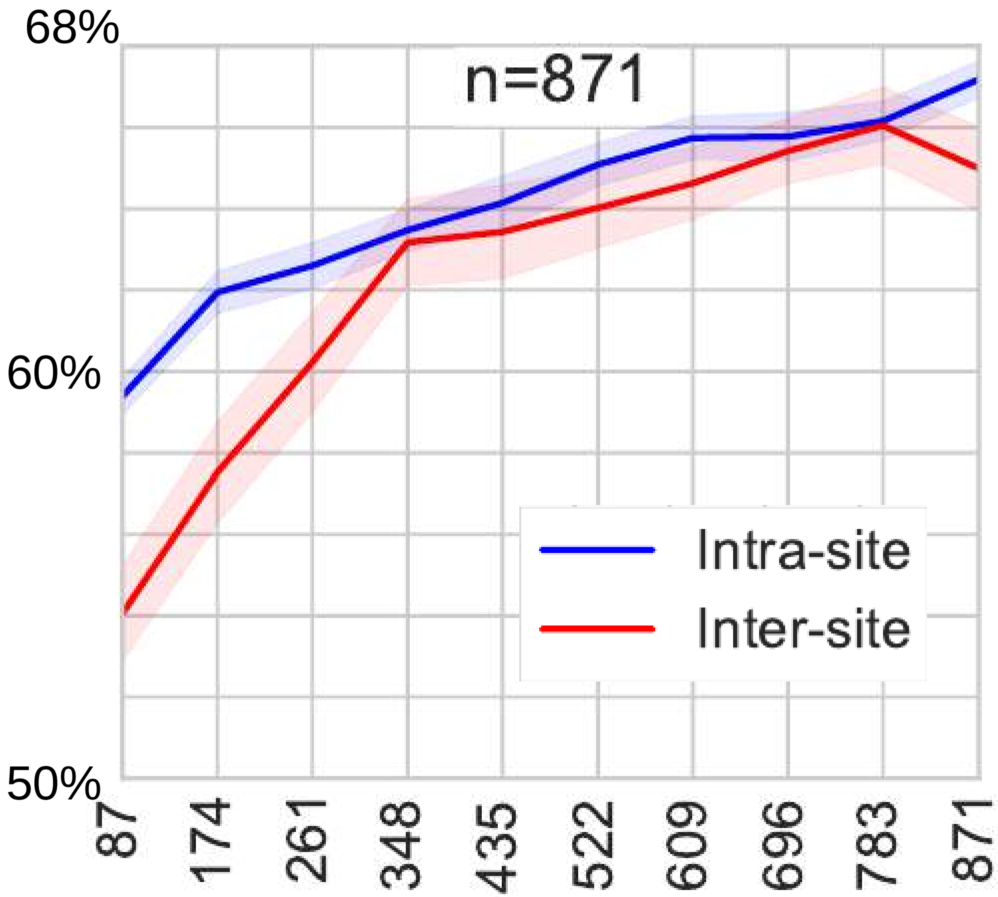


(effect size estimated by bootstrap)



Getting more data to feed learning machines

- Multi-sites cohort



classification accuracy **of the ABIDE dataset**

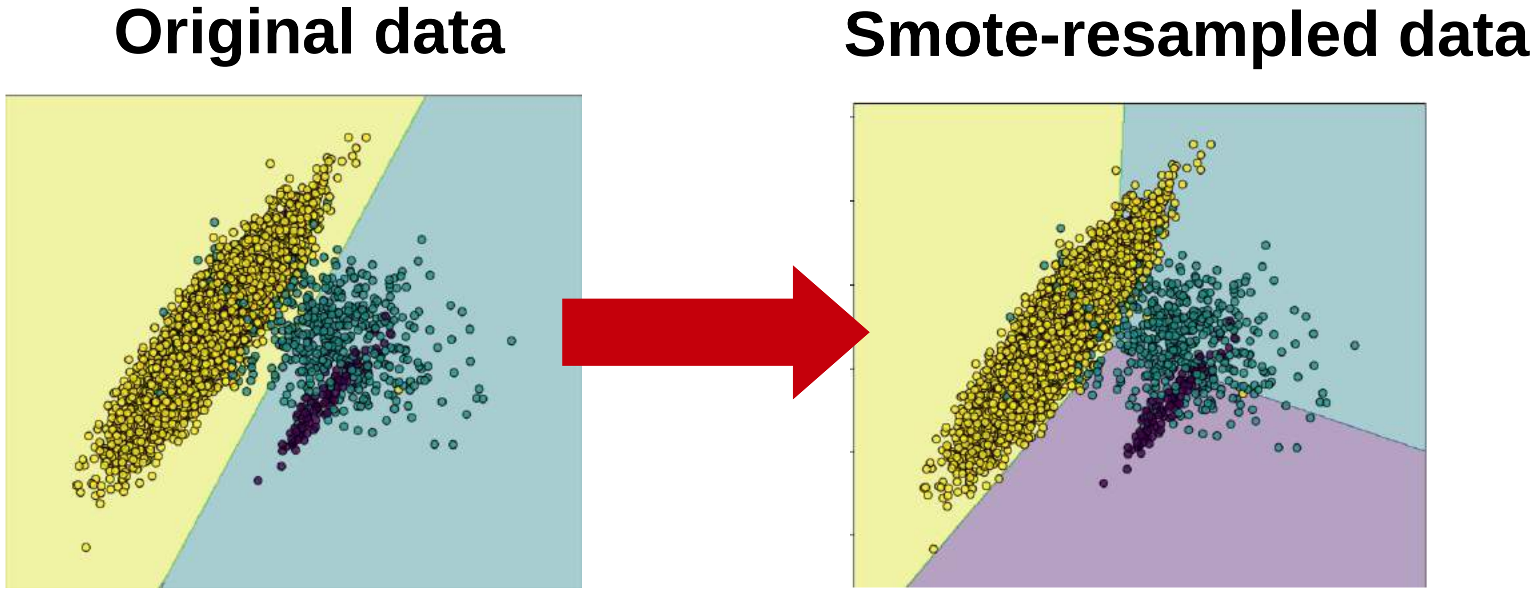
- **n=871 subjects**
- **17 acquisition centers**

Increasing curves → more subjects in the training set improves prediction accuracy.

[Abraham et al. Nimg 2016]

Getting more data to feed learning machines

- covariate shift
- class imbalance
- long-tailed distribution of labels

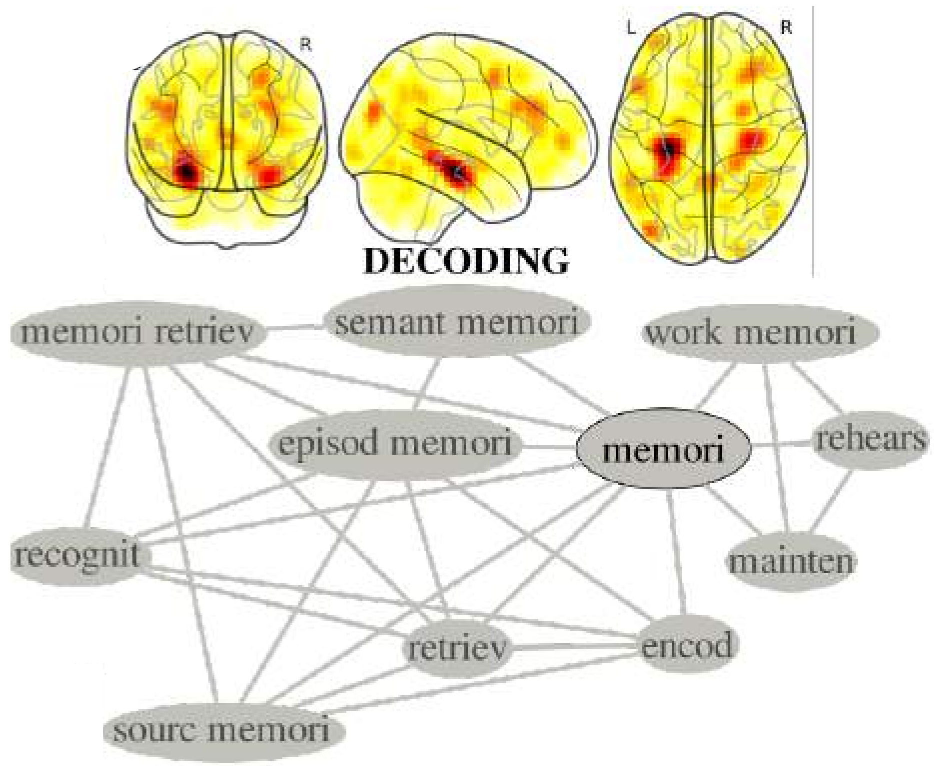


http://contrib.scikit-learn.org/imbalanced-learn

Getting more data to feed learning machines

- The cost is data annotation
- Try to glean concepts organization from the literature + learn association between terms and imaging structures

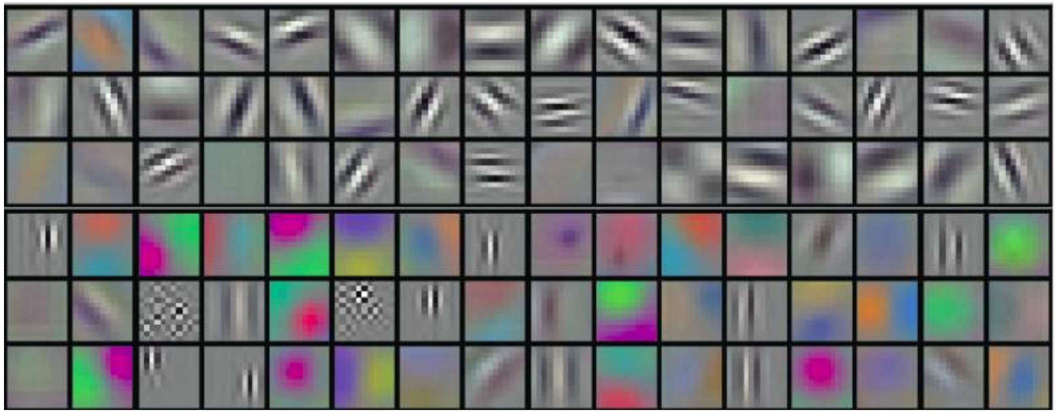
"Memory"



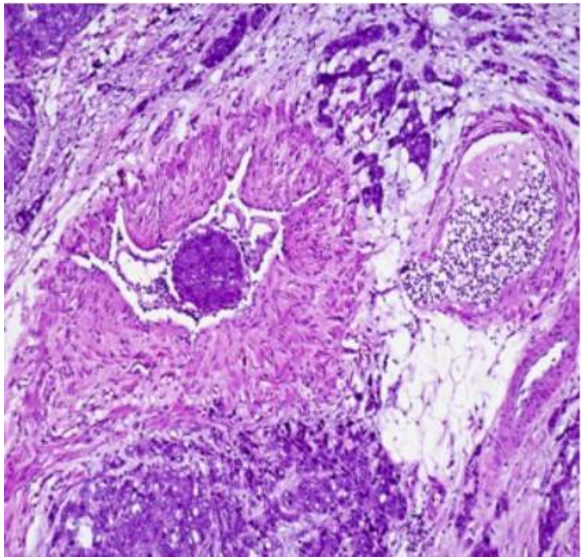
[Dockes et al MICCAI 2018 accept]

Overcoming the lack of data

- Transfer learning



ImageNet features



Breast cancer diagnosis

- Data augmentation: transformations, rescaling, resampling, deformations.

ML on large image datasets

HCP mailing list, Jan 19th, 2015

“Has anyone on the ML run group-wise analysis on the HCP resting state data, and if so what tools did you use?

I am having memory issues when running more than 10 subjects and I was wondering if anyone has a way of getting around the large memory requirements when concatenating in time.”

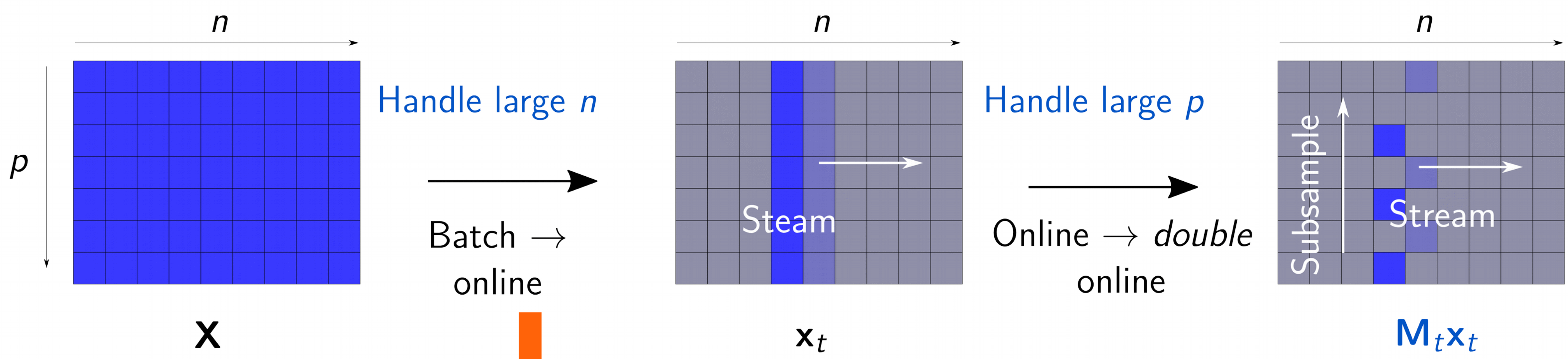
Working on huge data matrices

- Human Connectome project
- $n=2.10^6$, $p=2.10^5$, **2TB** of data
- Task: segmentation into regions
- Online dictionary learning [Mairal et al. ICML 2009]
- How to go faster ?
  - Work on batches of images **and** voxels

[Mensch et al. ICML 2016, IEEE TSP 2018]

Stochastic gradient approaches

http://amensch.fr/research/2016/06/10/modl.html

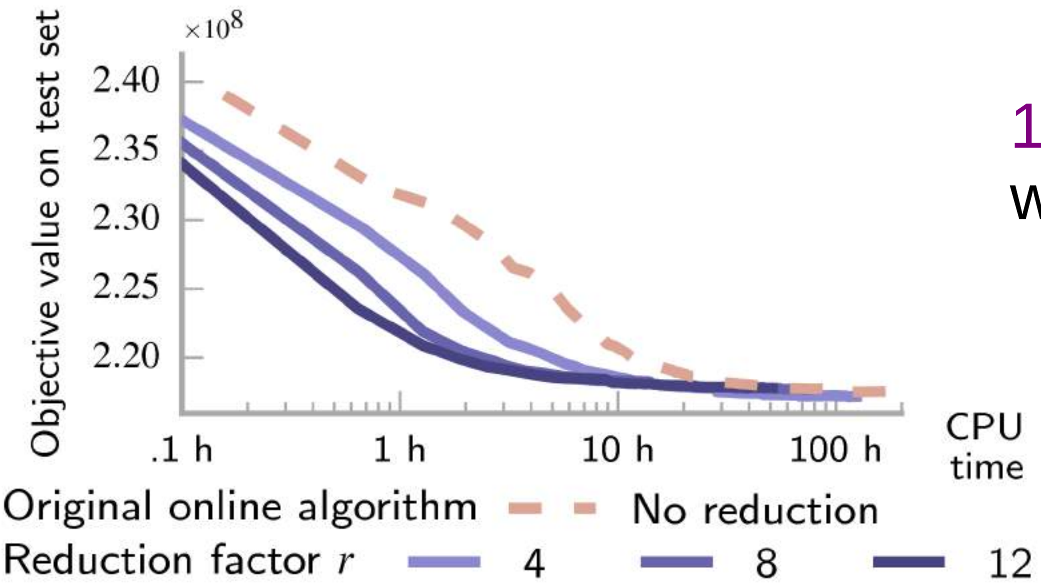


$$\boldsymbol{\alpha}_t(\mathbf{D}) = \mathrm{argmin}_{\mathbf{A}\in\mathbb{R}^{k\times n}} \|\mathbf{x}_t - \mathbf{D}_{t-1}\boldsymbol{\alpha}_t\|_F^2 + \lambda\Omega(\boldsymbol{\alpha}_t)$$

$$\mathbf{D}_t = \mathrm{argmin}_{\mathbf{D}\in\mathcal{C}} \sum_{i=1}^{t} \|\mathbf{x}_i - \mathbf{D}\boldsymbol{\alpha}_i\|_F^2$$

$$\boldsymbol{\alpha}_t(\mathbf{D}) = \mathrm{argmin}_{\mathbf{A}\in\mathbb{R}^{k\times n}} \|\mathbf{M}_t(\mathbf{x}_t - \mathbf{D}_{t-1}\boldsymbol{\alpha}_t)\|_F^2 + \lambda\frac{s}{p}\Omega(\boldsymbol{\alpha})$$

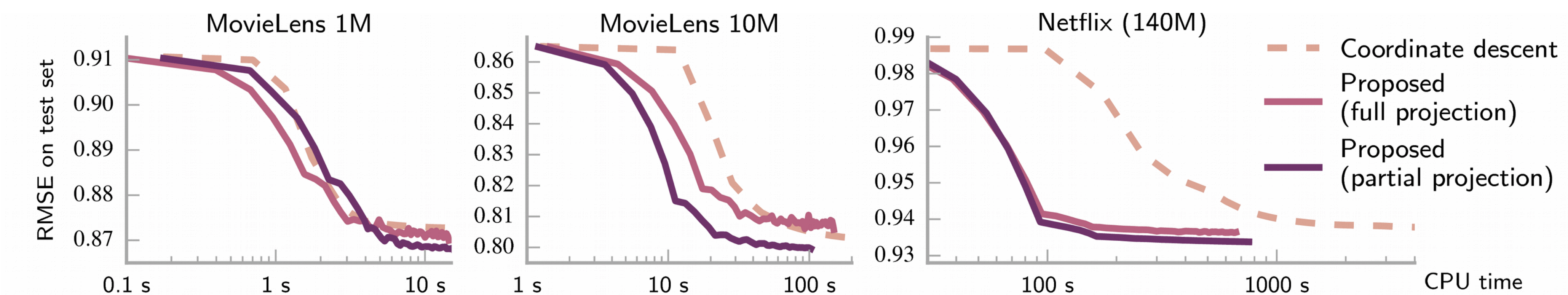
Stochastic gradient approaches



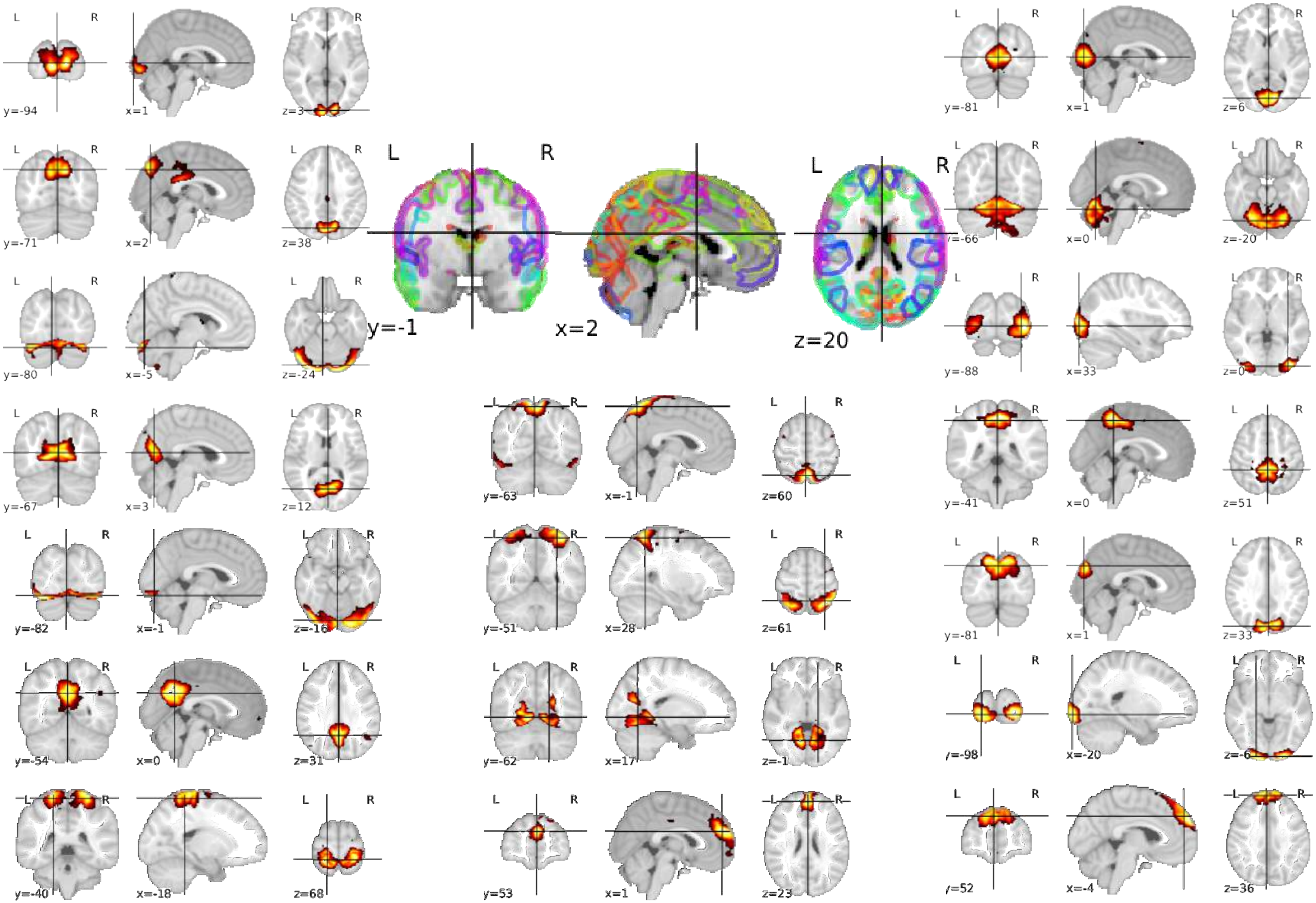
10-fold gain in CPU time
without loss in accuracy

[Mensch et al. ICML 2016, IEEE TSP 2018]

Can be used for recommender systems

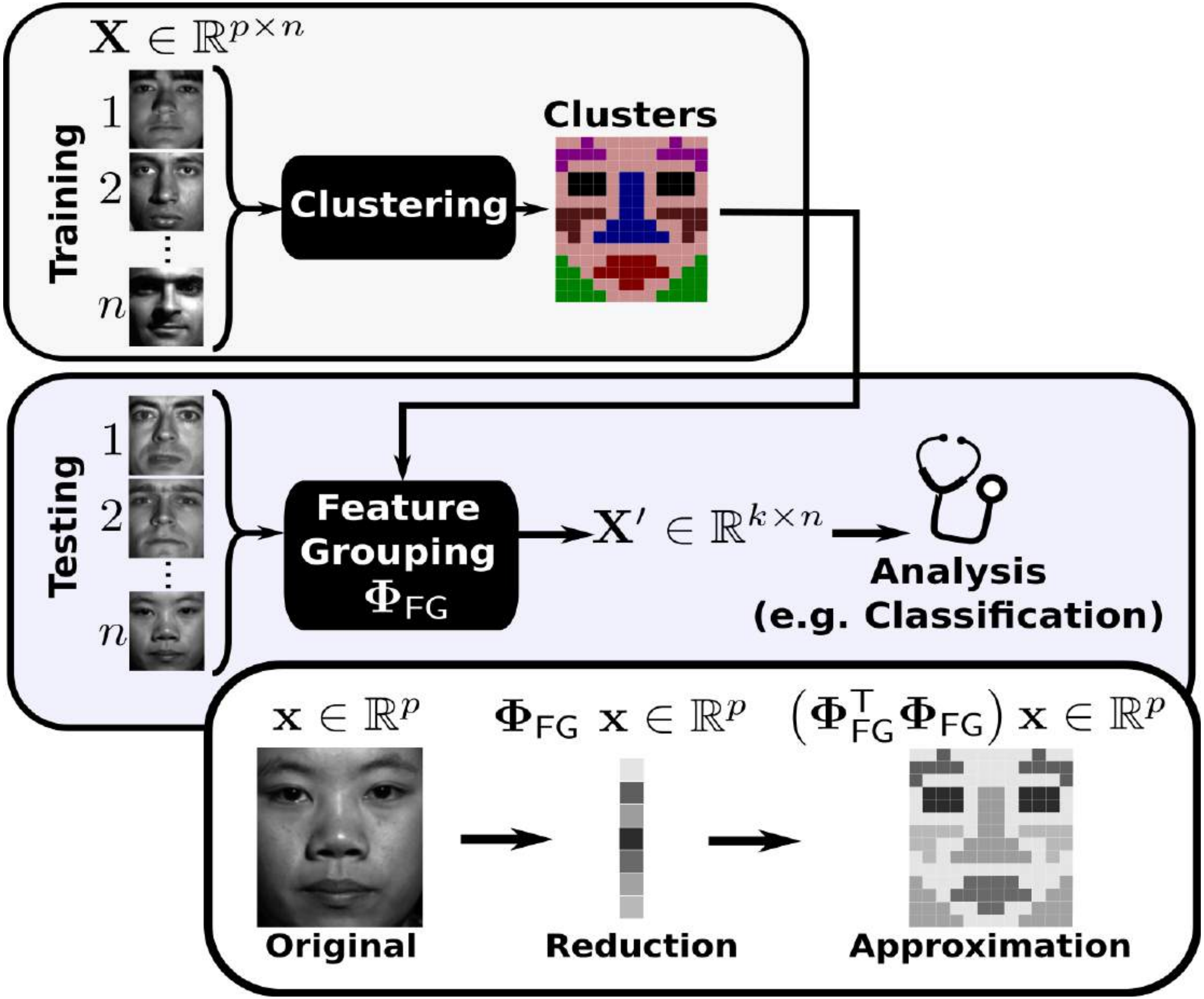


Resulting brain atlas

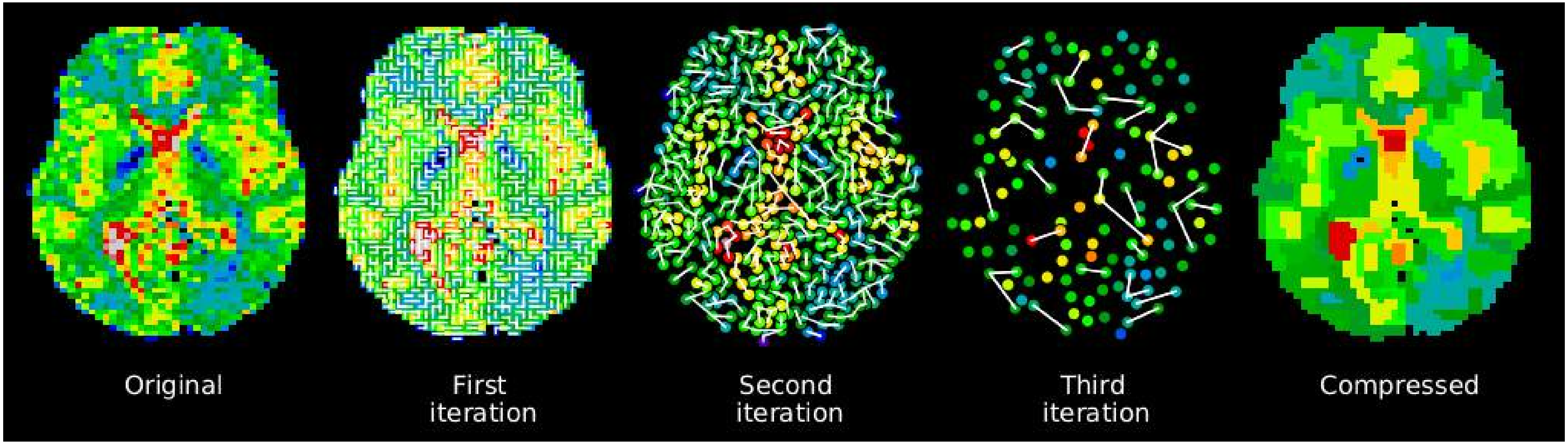


[Mensch et al. ICML 2016 IEEE TSP 2018]

Compression by feature grouping



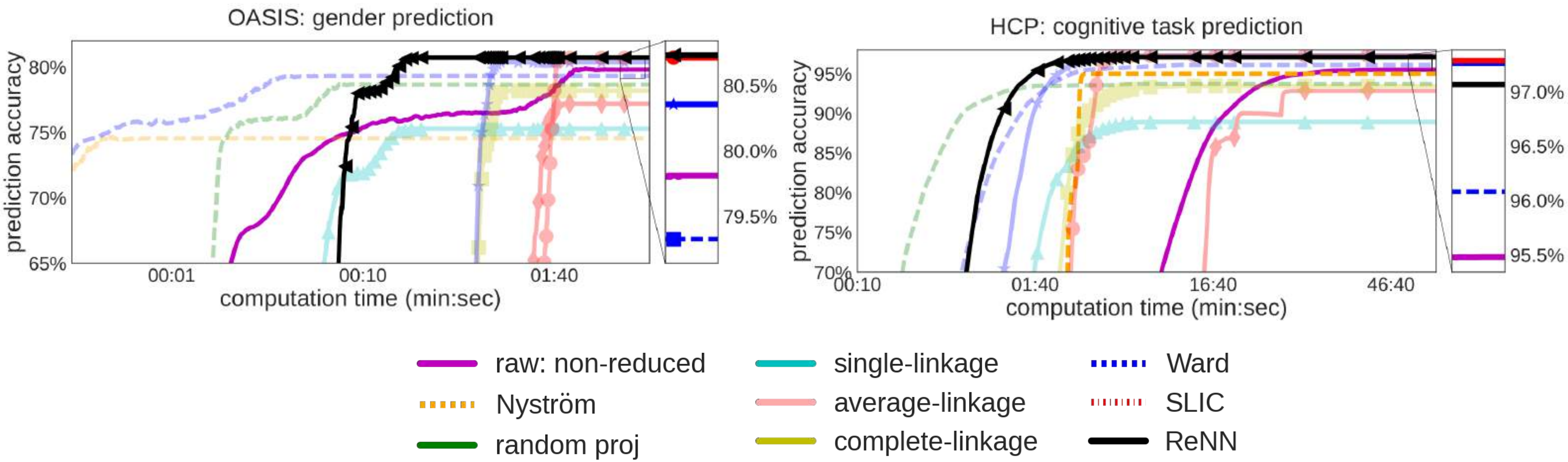
Recursive Neighbour Agglomeration



Based on local decisions = fast (linear time) – avoid percolation

[Thirion et al. Stamlins 2015, Idrobo et PAMI in press]

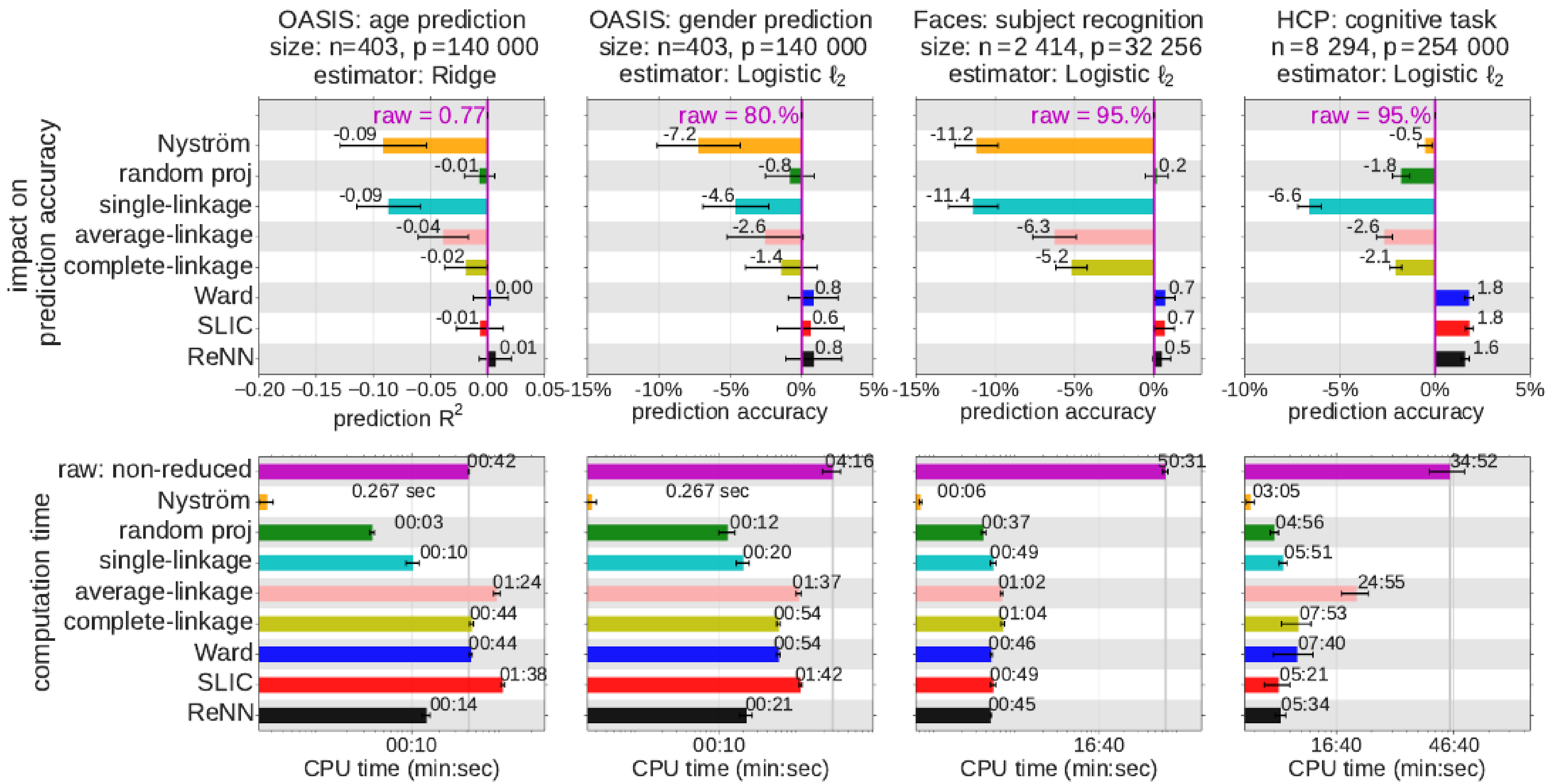
Effect on data analysis tasks



- Impressive speed-up and increased accuracy with respect to non-compressed representation
  - Clustering has a denoising effect

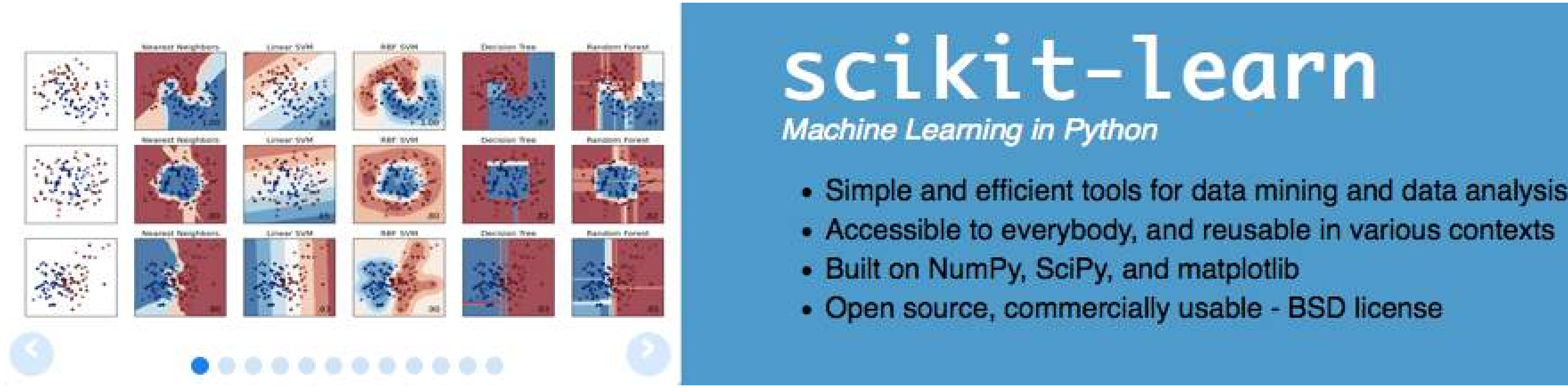
[Hoyos Idrobo IEEE PAMI in press]

More results



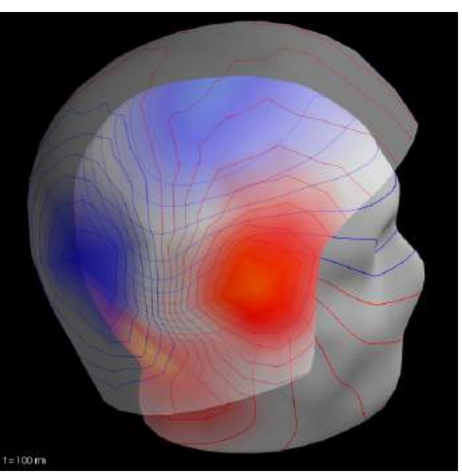
[Hoyos Idrobo IEEE PAMI in press]

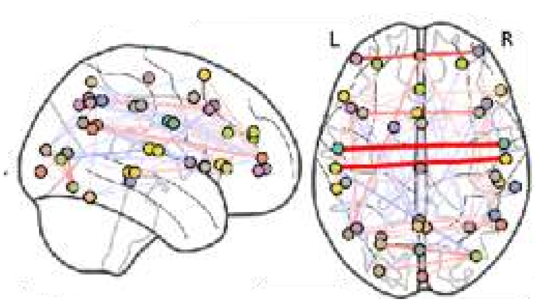
Software

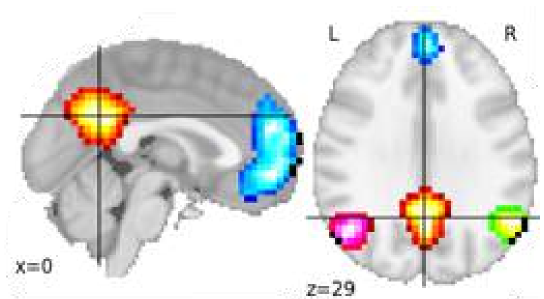


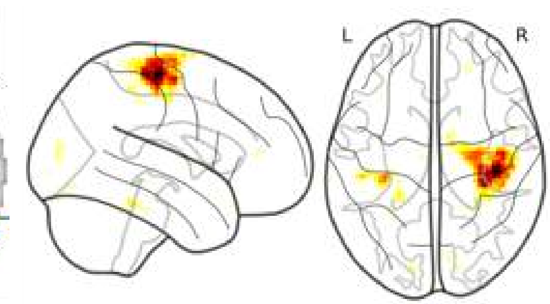
- Python OSS, community development
- scikit learn: all types of shallow machine learning
- MNE, nilearn: brain imaging applications

MEG + EEG ANALYSIS & VISUALIZATION









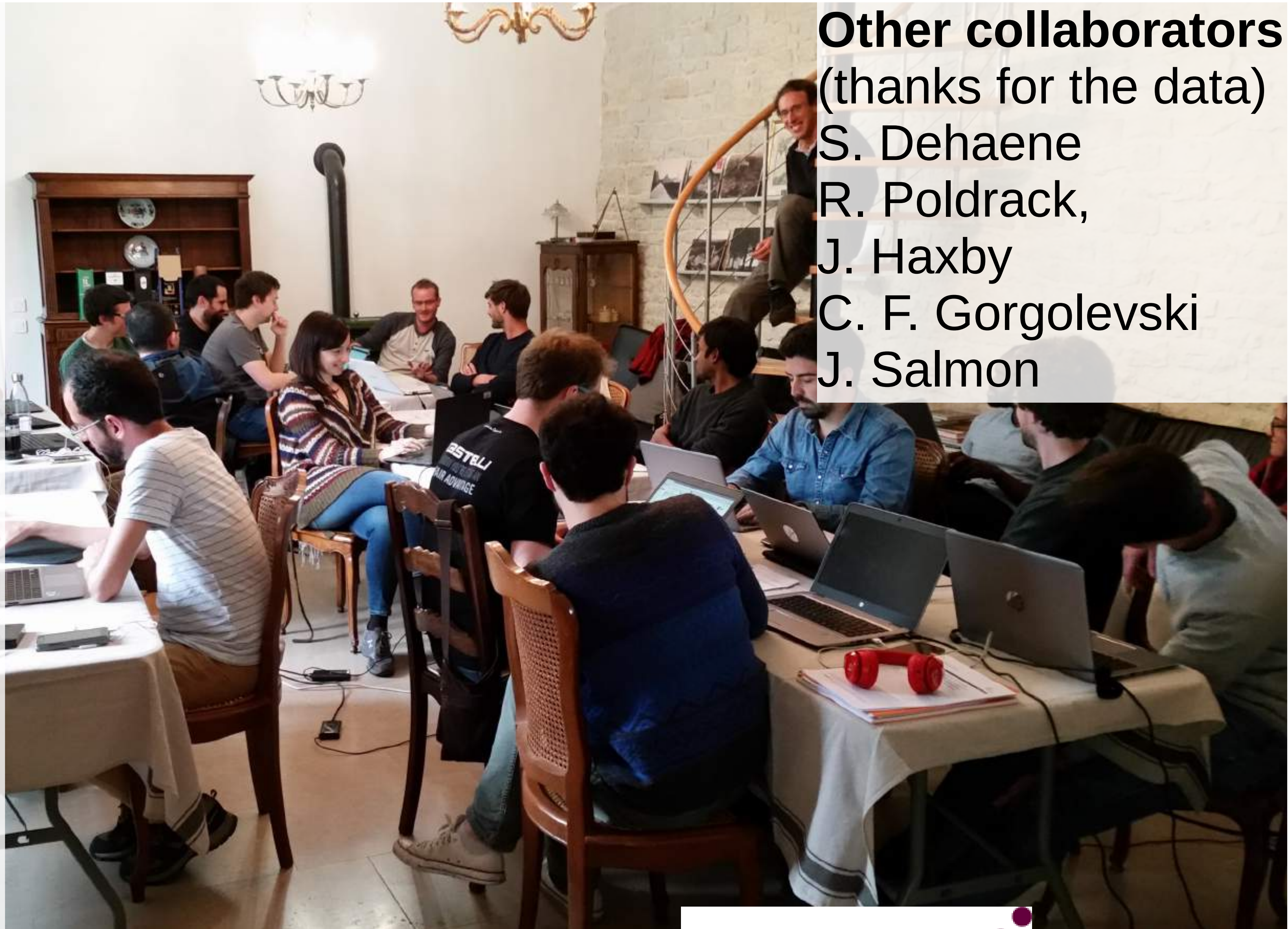
Conclusion

- Dataset increase: Importance of data sharing
- Deal with confounds and covariate shifts
- Handling of missing data, inconsistent annotations
- Explicability of models

Acknowledgement

Parietal

G. Varoquaux,
A. Gramfort,
P. Ciuciu,
D. Wassermann,
D. Engemann,
A. Manoel,
D. Chyzhyk
A.L. Grilo Pinho,
E. Dohmatob,
A. Mensch,
J.A. Chevalier,
A. Hoyos idrobo,
D. Bzdok,
J. Dockès,
P. Cerda,
C. Lazarus
D. La Rocca
G. Lemaitre
L. El Gueddari
O. Grisel
M. Massias
P. Ablin
H. Janati
J. Massich
K. Dadi
C. Petitot



Other collaborators
(thanks for the data)
S. Dehaene
R. Poldrack,
J. Haxby
C. F. Gorgolevski
J. Salmon

Human Brain Project

université PARIS-SACLAY

ANR