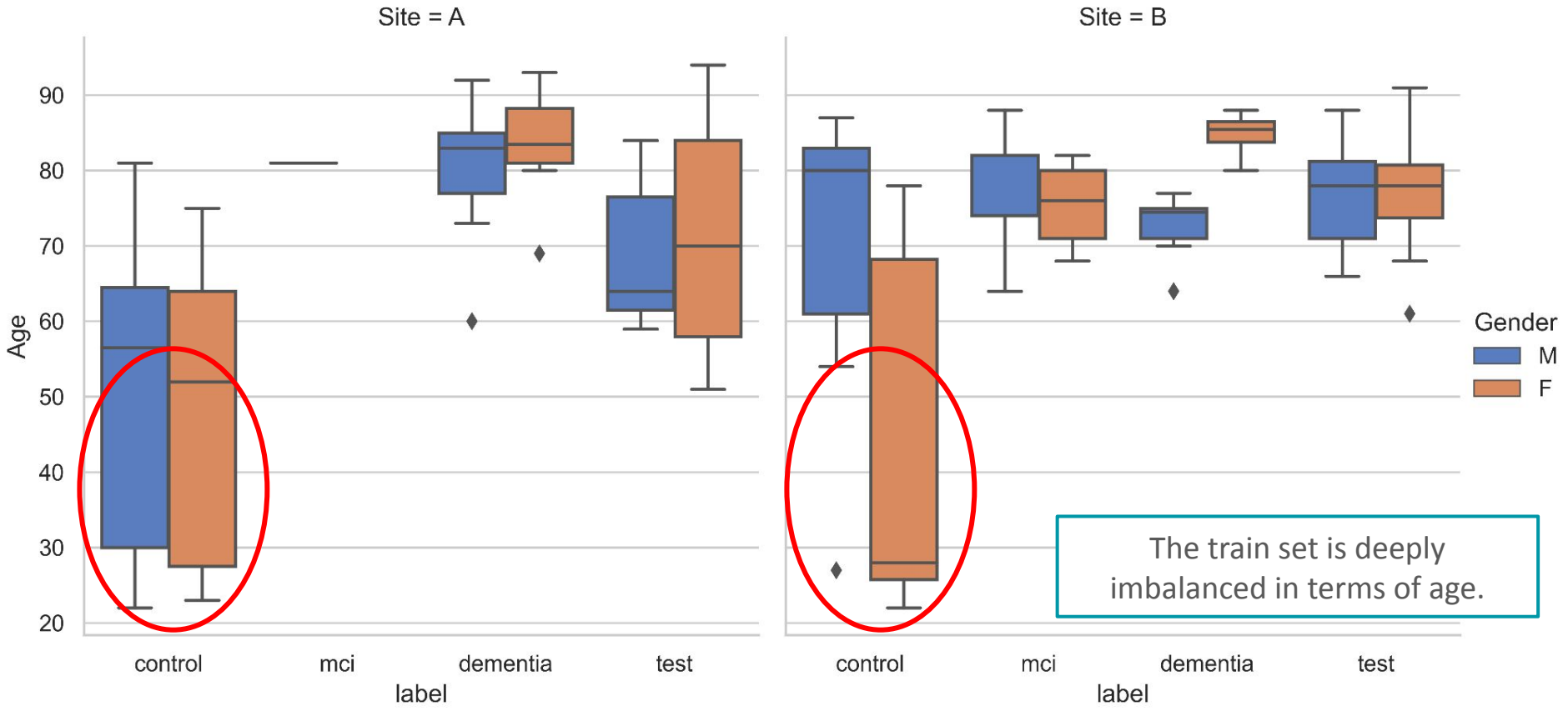
Data Analysis Competition: Dementia screening challenge

Apolline MELLOT - Benoît MALEZIEUX - Cédric ALLAIN

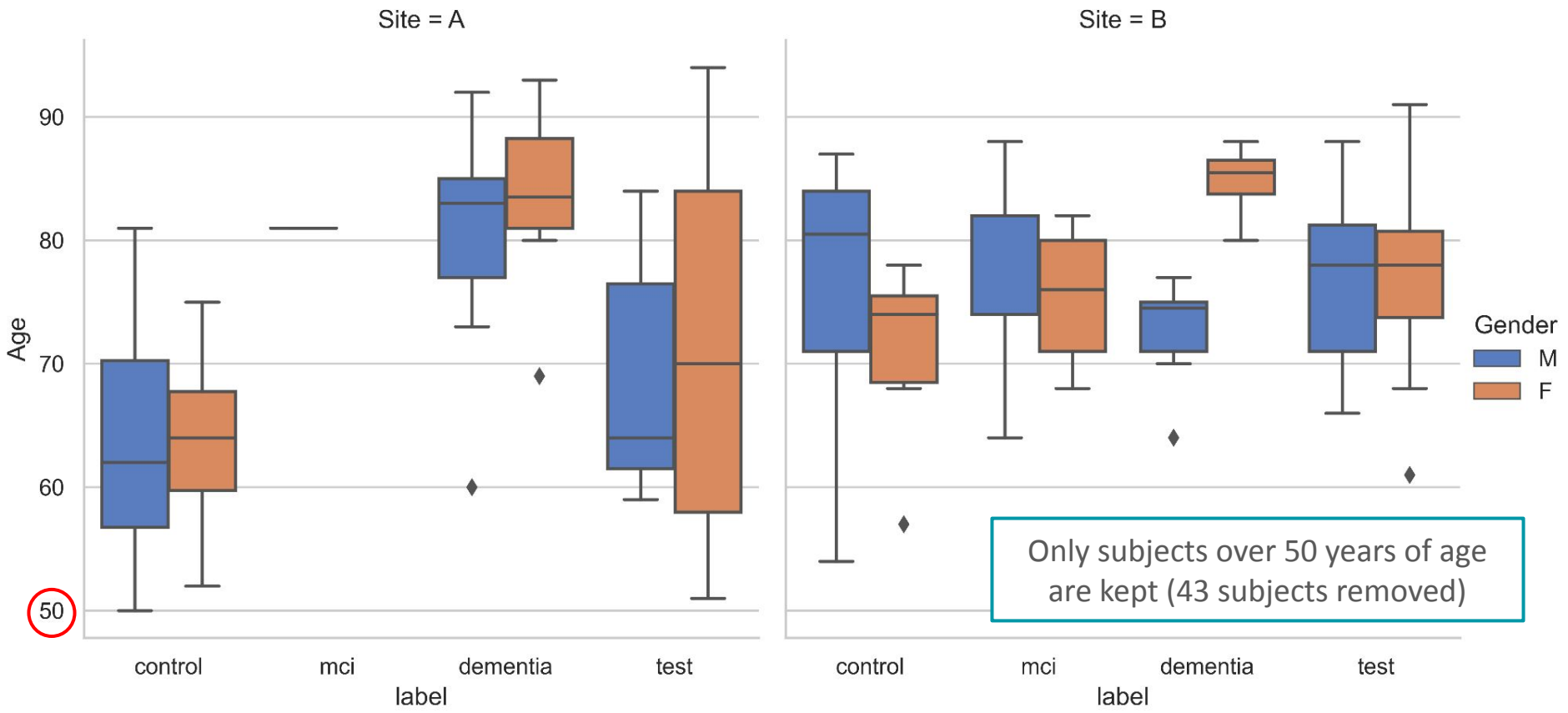
Université Paris-Saclay, Inria, CEA, Palaiseau, France

https://github.com/apmellot/Dementia_screening_challenge_biomag_2022

Age distribution across sites



Age distribution across sites



Classification pipeline

Preprocessing:

MNE-BIDS pipeline : https://mne.tools/mne-bids-pipeline/index.html

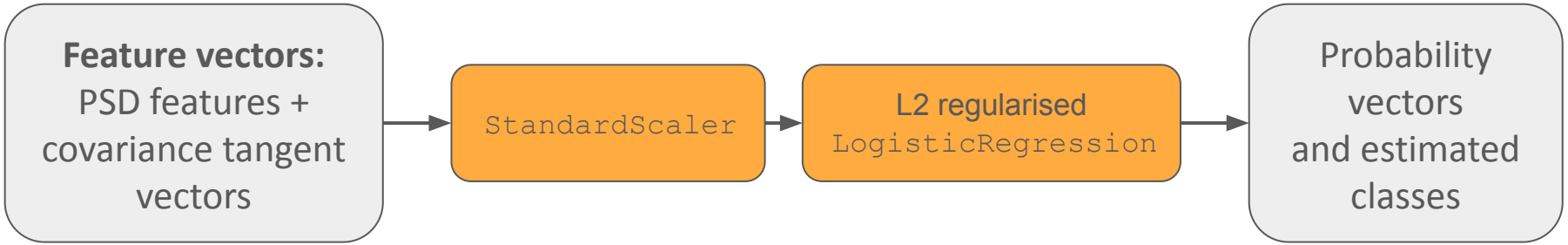
poster IT-75

➔ Resampling at 200 Hz

➔ Bandpass between 0.1 Hz and 49 Hz

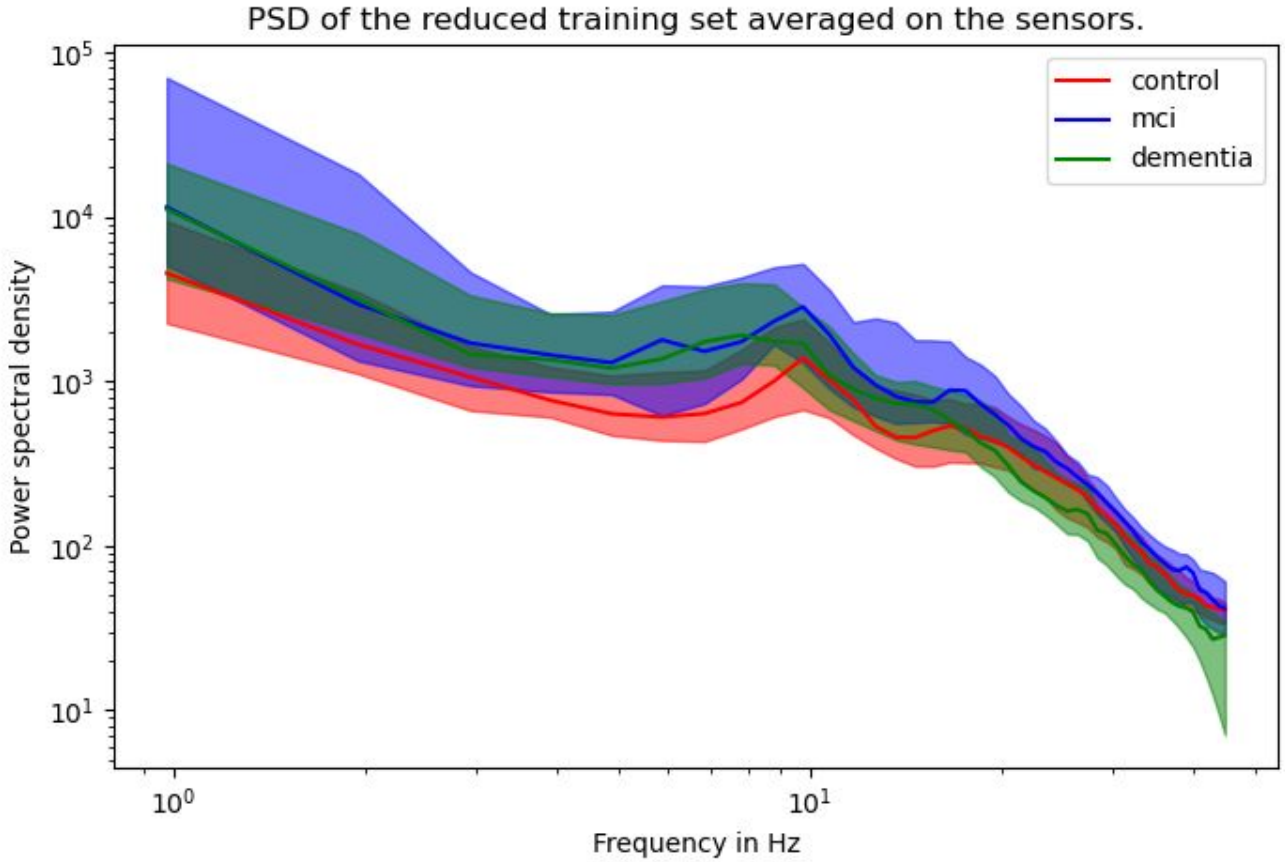
➔ 10 second epochs without overlap

Classification pipeline:



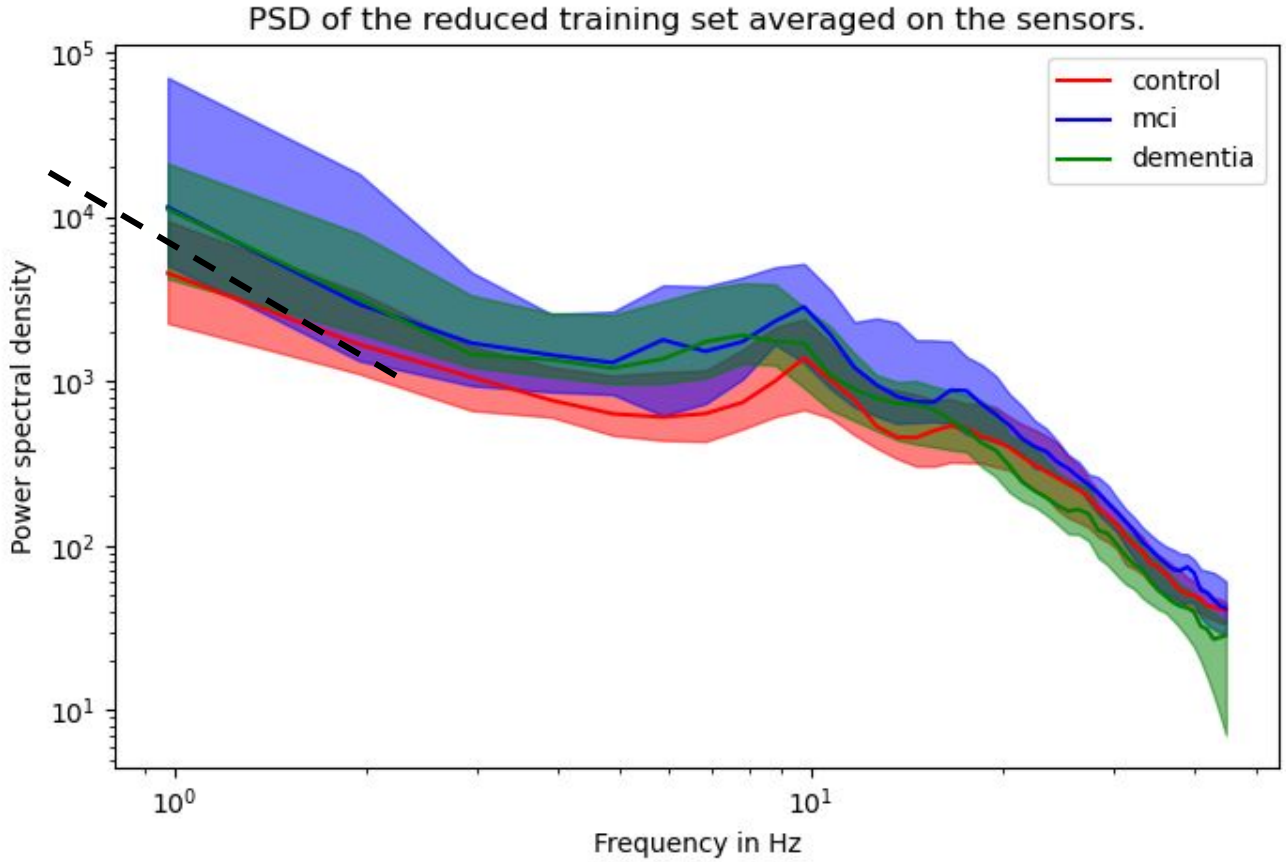
[Engemann et al. 2020, 2022]

Features extracted from the Power Spectral Densities:



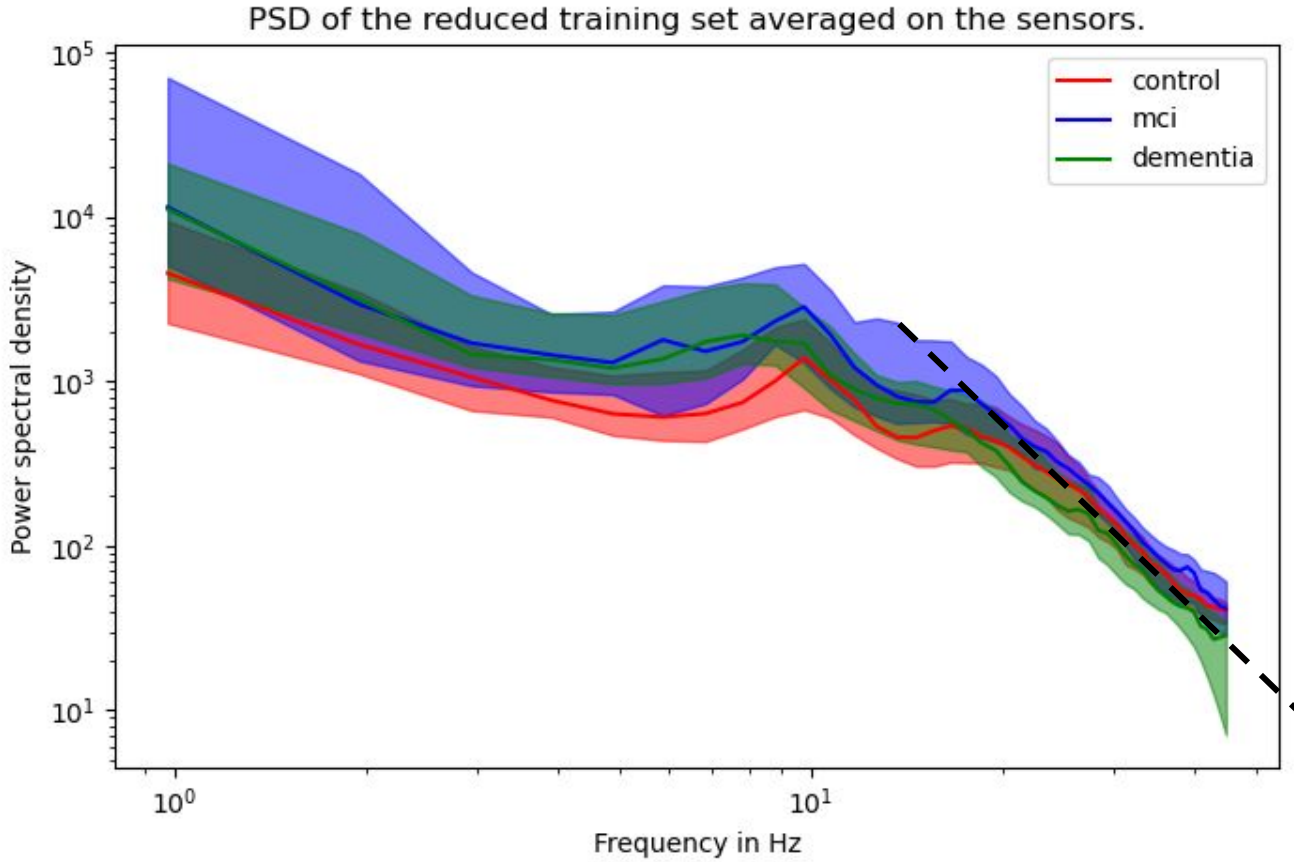
Features extracted from the Power Spectral Densities:

➔ 1/f slopes between 0.1 to 1.5 Hz (low)



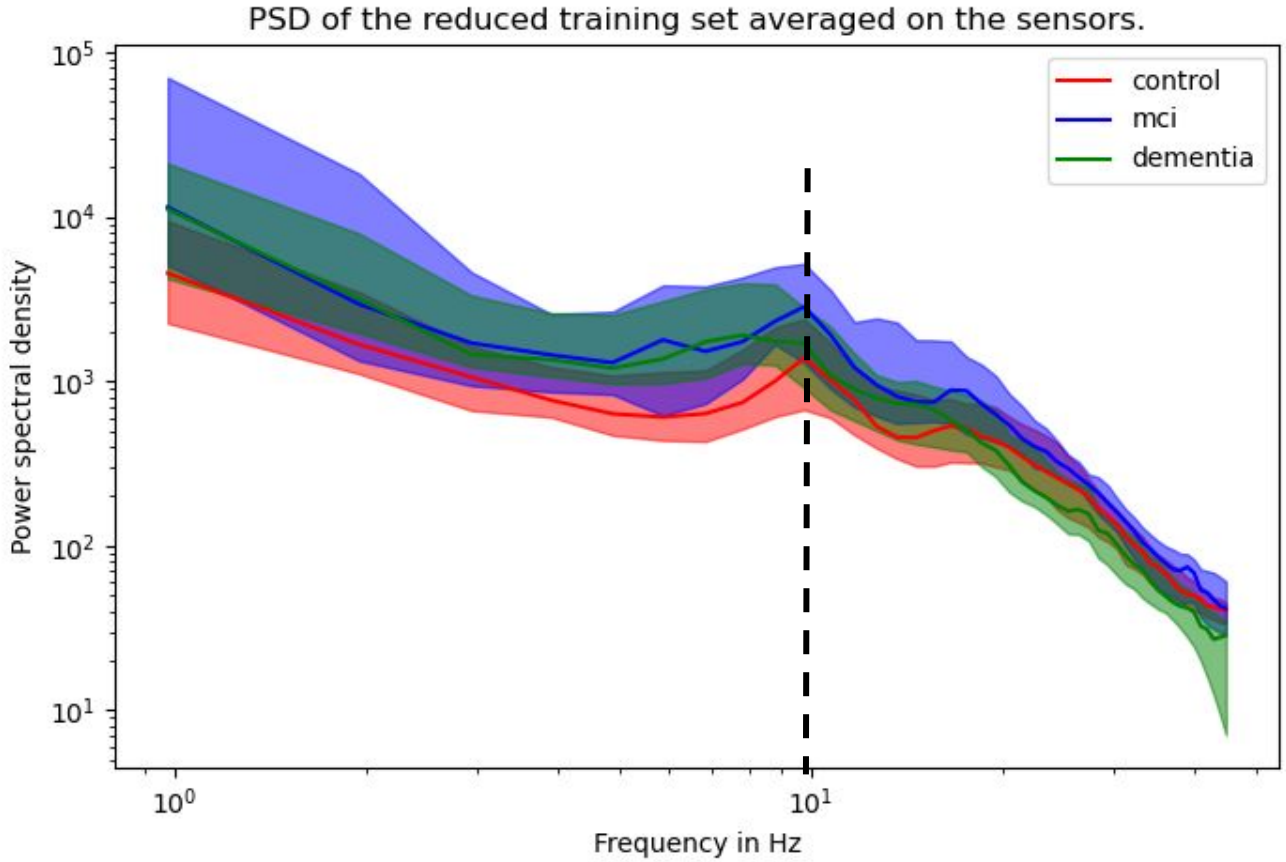
Features extracted from the Power Spectral Densities:

- ➔ 1/f slopes between 0.1 to 1.5 Hz (low)
- ➔ 1/f slopes between 35 to 49 Hz (beta high)



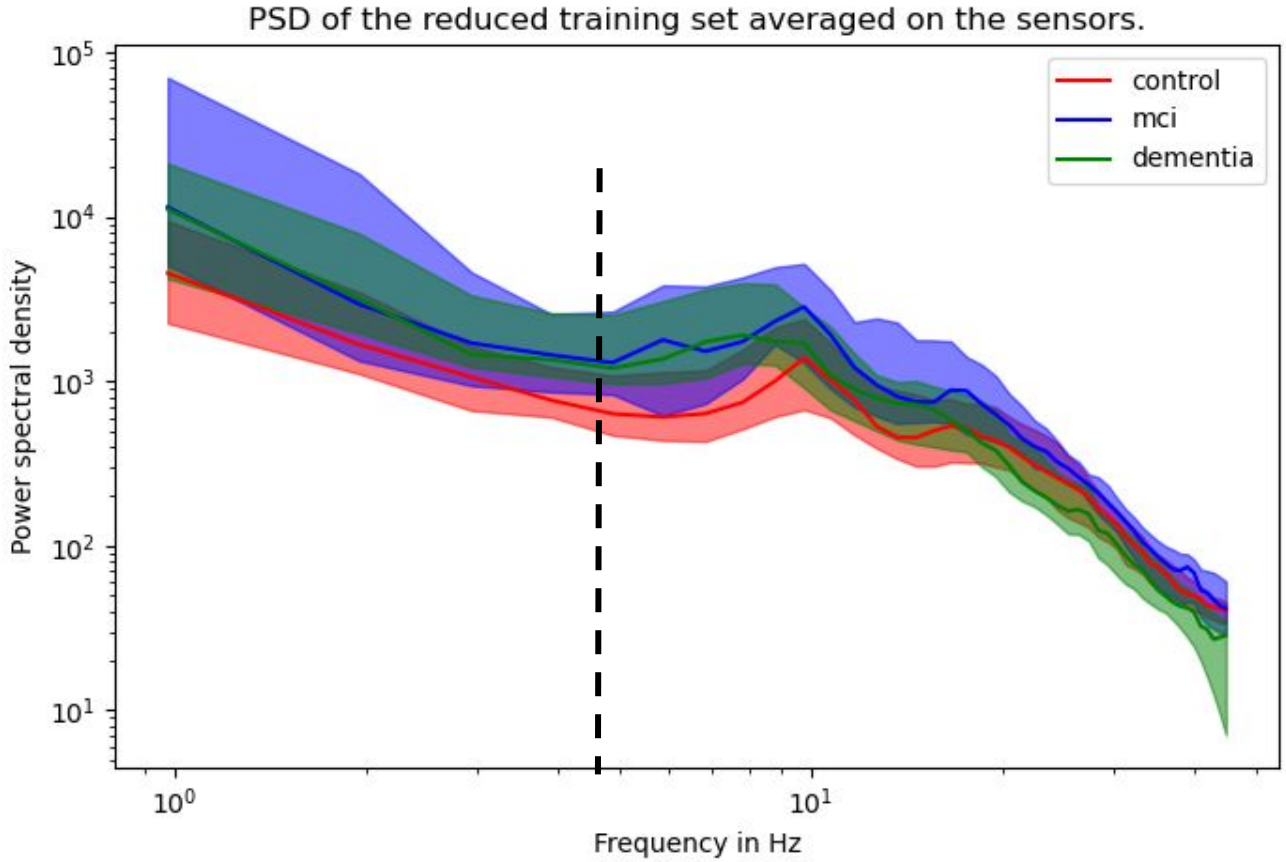
Features extracted from the Power Spectral Densities:

- ➔ 1/f slopes between 0.1 to 1.5 Hz (low)
- ➔ 1/f slopes between 35 to 49 Hz (beta high)
- ➔ Alpha peak frequency



Features extracted from the Power Spectral Densities:

- ➔ 1/f slopes between 0.1 to 1.5 Hz (low)
- ➔ 1/f slopes between 35 to 49 Hz (beta high)
- ➔ Alpha peak frequency
- ➔ Median power frequency



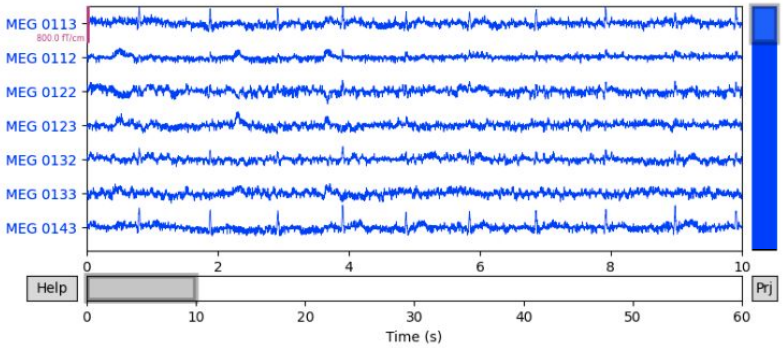
Feature vectors extracted from covariances:

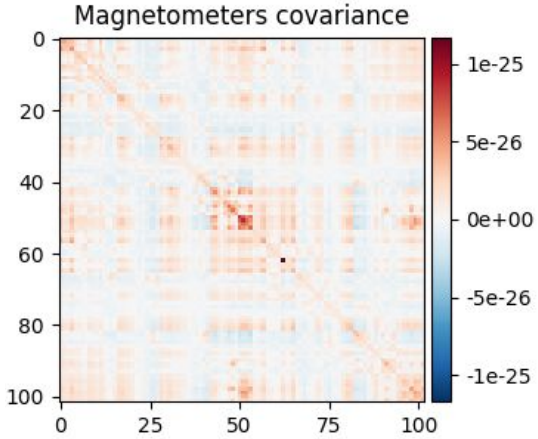
Prerocessed M/EEG signal: Covariances:

$$X(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix} \in \mathbb{R}^{n \times T_S} \longrightarrow C = \frac{1}{T_s - 1} X(t) X(t)^T \in \mathbb{R}^{n \times n}$$

where *n* is the number electrodes and T_s the number of sampled time points

Raw signals





https://mne.tools/stable/index.html

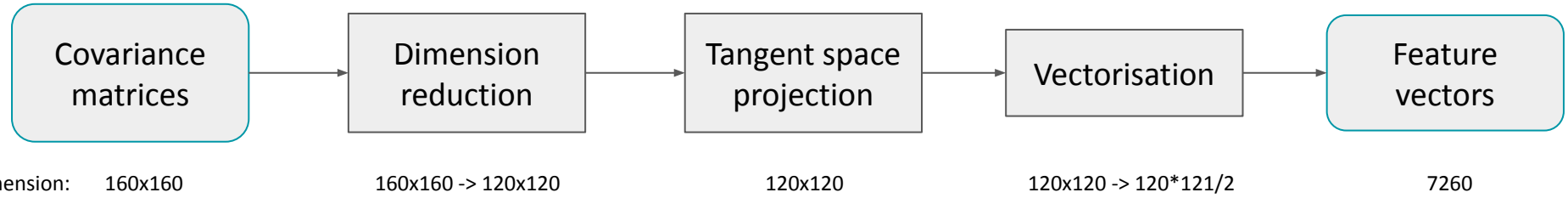
Feature vectors extracted from covariances:

Definition of frequency bands

Name	low	δ	θ	α	β_{low}	β_{mid}	β_{high}
Range (Hz)	0.1 - 1	1 - 4	4 - 8	8 - 15	15 - 26	26 - 35	35 - 49

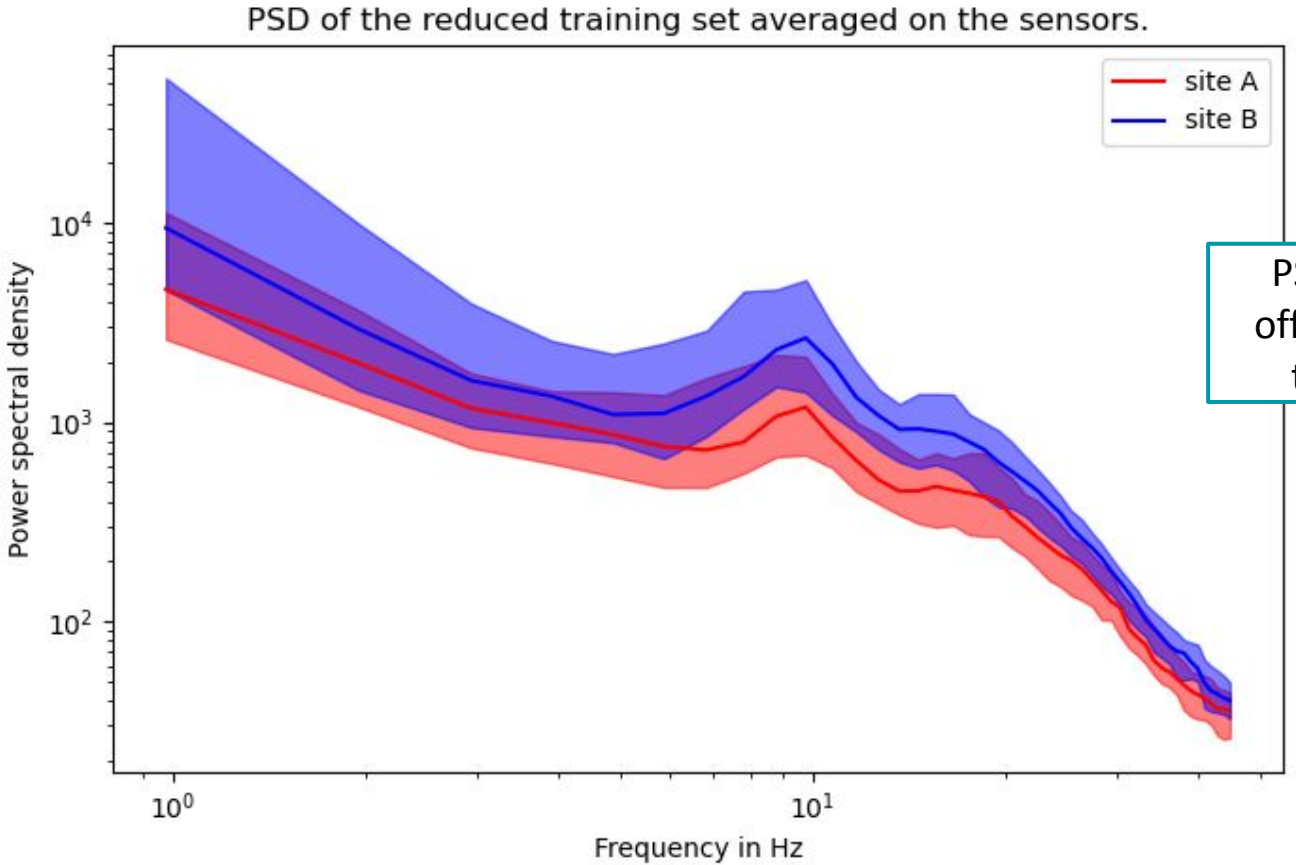
➔ One covariance matrix per frequency band averaged over all epochs

Pipeline:



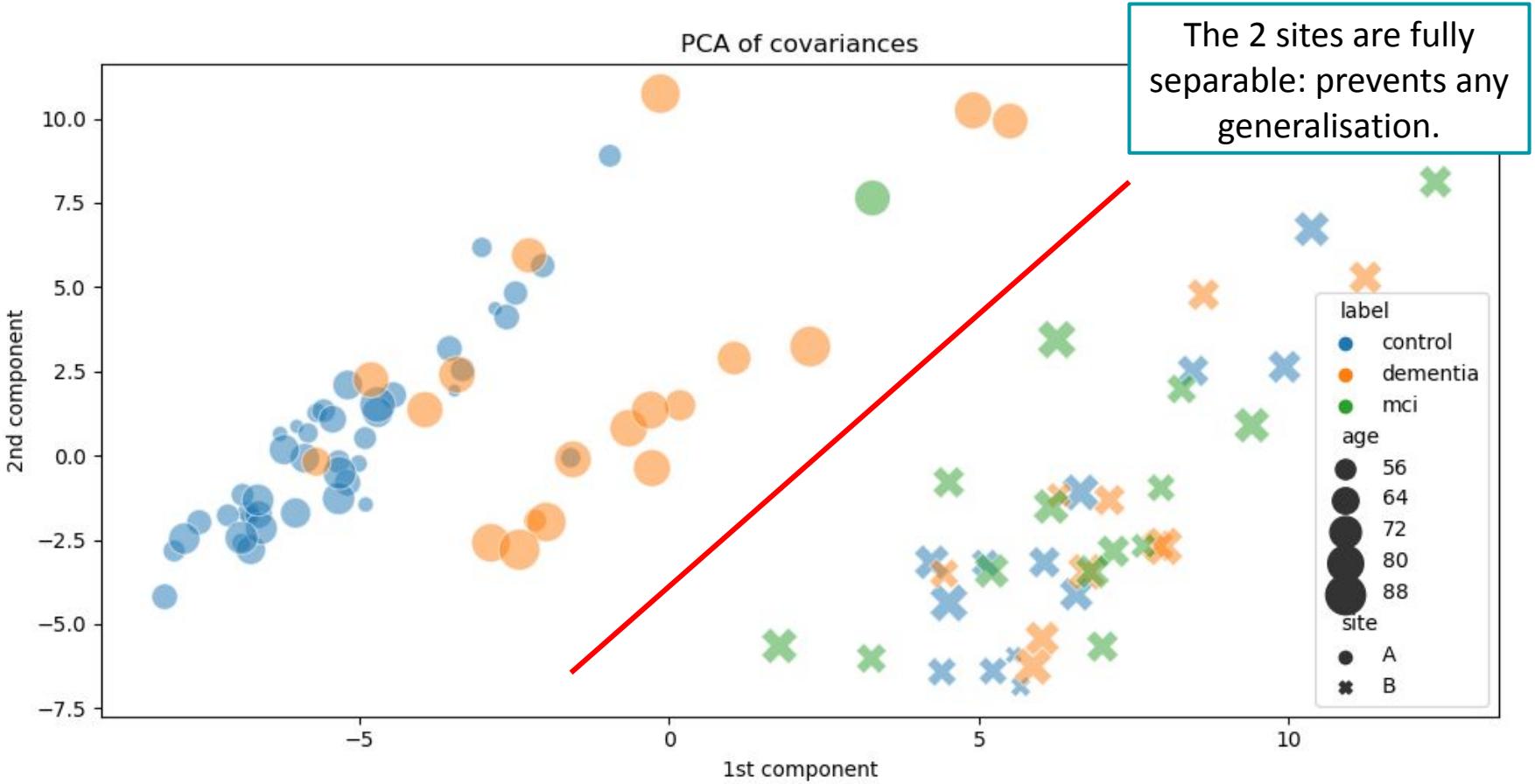
[Sabbagh et al. 2019, 2020] [Engemann et al. 2022]

Remaining issue: recording site



PSD intensity offset between the 2 sites.

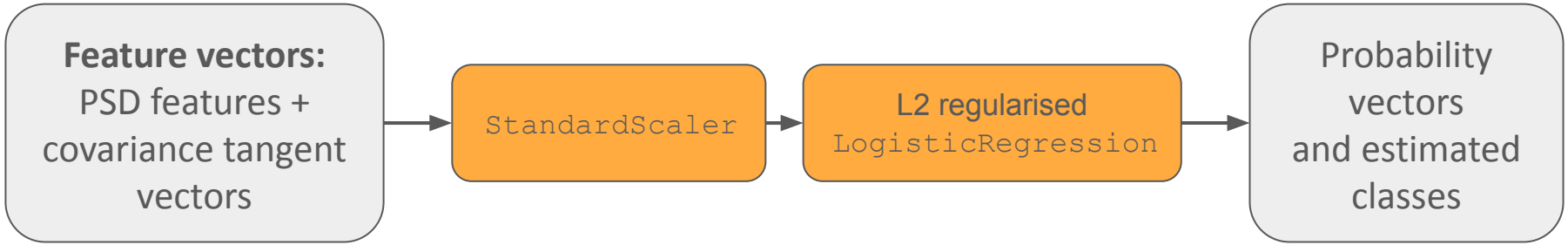
Remaining issue: recording site



The 2 sites are fully separable: prevents any generalisation.

Results on training set

Classification pipeline:



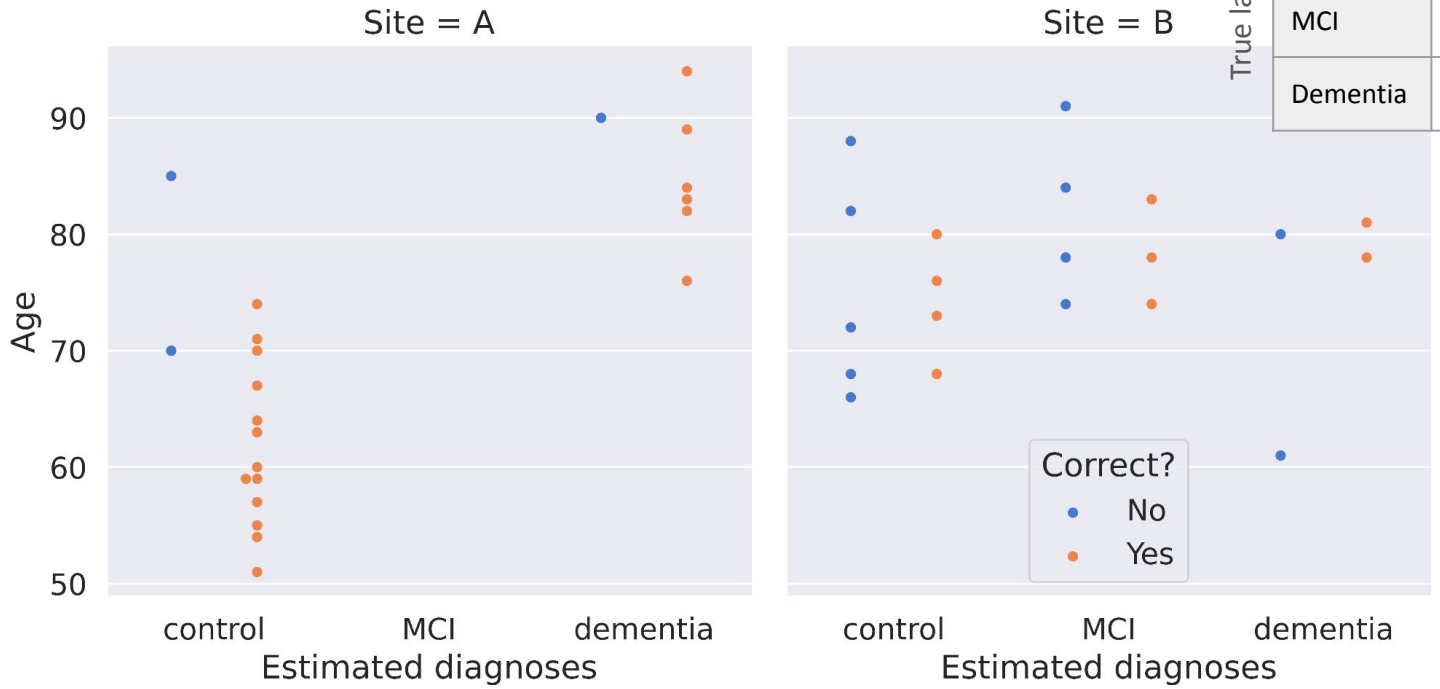
Results on the training set:

➔ `StratifiedShuffleSplit` cross-validation from scikit-learn with 20 splits and `test_size = 0.2`

Mean accuracy: 71.9% ± 8.3%

Mean balanced accuracy: 64.3% ± 9.6%

Results on test set



True label \ Predicted label	Control	MCI	Dementia
Control	17	2	1
MCI	3	3	2
Dementia	4	2	8

Confusion matrix on test set

➔ High accuracy for unseen data from site A (86.3%) but low accuracy for site B (45%).

Acknowledgment

- Organisers of this competition.
- Alexandre GRAMFORT, Denis ENGEMANN for encouraging us to work on this challenge and for sharing their machine learning expertise.
- The Inria MIND Team.
- ANR funding AI-cog (ANR-20-IADJ-0002)

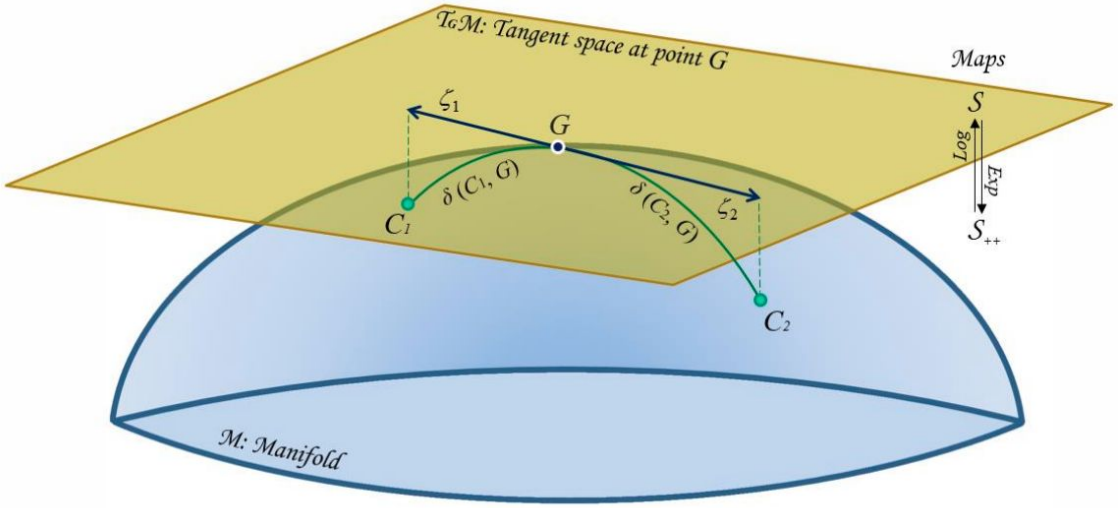
Results on test set

	Recall	Precision	F-score
Control	85%	71%	78%
MCI	38%	43%	40%
Dementia	57%	73%	65%
	mean recall	mean precision	mean F-score
	60%	62%	61%

Overall accuracy: 67%

Riemannian geometry concepts

Log of a matrix: $M = U\Delta U^\top \mapsto \log(M) = U\log(\Delta)U^\top$



Log map:

$$\zeta_1 = \mathrm{Log}_G(C_1) = G^{\frac{1}{2}}\log(G^{-\frac{1}{2}}C_1G^{-\frac{1}{2}})G^{\frac{1}{2}}$$

Exp map:

$$C_1 = \mathrm{Exp}_G(\zeta_1) = G^{\frac{1}{2}}\exp(G^{-\frac{1}{2}}\zeta_1G^{-\frac{1}{2}})G^{\frac{1}{2}}$$

Riemannian distance:

$$\delta_R(C_1, C_2) = ||\log(C_1^{-1}C_2)||_F$$

Affine invariance: $\delta_R(C_1, C_2) = \delta_R(AC_1A^\top, AC_2A^\top)$

Vector representation: $z_1 = \mathrm{uvec}(\zeta_1 \odot Q) \in \mathbb{R}^{\frac{P(P+1)}{2}}$ with Q a matrix holding 1 on the diagonal elements and $\sqrt{2}$ elsewhere.

[Barachant et al. 2011] [Congedo et al. 2017]