Functional connectivity analysis on ABIDE

PIERNICOLA OLIVA
UNIVERSITY OF SASSARI & INFN CAGLIARI
OLIVA@UNISS.IT

ABIDE

Autism spectrum disorder (ASD) is characterized by qualitative impairment in social reciprocity, and by repetitive, restricted, and stereotyped behaviors/interests.

The small advances in research are certainly also due to the complexity and heterogeneity of ASD

To face these challenges, large-scale samples are essential, but single laboratories cannot obtain sufficiently large datasets to reveal the brain mechanisms underlying ASD.

The Autism Brain Imaging Data Exchange (ABIDE) initiative has aggregated *functional* and *structural* brain imaging data collected from laboratories around the world to accelerate our understanding of the neural bases of autism.

The dataset contains **previously collected** MR imaging and phenotypic datasets made available for data sharing with the broader scientific community.

ABIDE

ABIDE
Autism Brain Imaging
Data Exchange

ABIDE I

- 17 international sites
  - 1112 subjects,
    - 539 ASD
    - 573 typical controls
  - ages 7-64 years

  - **ABIDE I preprocessed**
  - **Pipelines:**
    - Connectome Computation System (CCS)
    - Configurable Pipeline for the Analysis of Connectomes (CPAC)
    - Data Processing Assistant for Resting-State fMRI (DPARSF)
    - Neuroimaging Analysis Kit (NIAK)
  - **ROIs:**
    - Automated Anatomical Labeling (AAL):
    - Eickhoff-Zilles (EZ)
    - Harvard-Oxford (HO):
    - Talaraich and Tournoux (TT):
    - Dosenbach 160:
    - Craddock 200 & Craddock 400

ABIDE II

- 19 international sites
    - 10charter institutions
    - 7new members
  - 1114 subjects,
    - 521 ASD
    - 593 typical controls
  - ages 5-64 years

“with greater phenotypic characterization, particularly in regard to measures of core ASD and associated symptoms”

38 individuals at two time points (1-4 year interval).

Preprocessed data are not available for ABIDE II

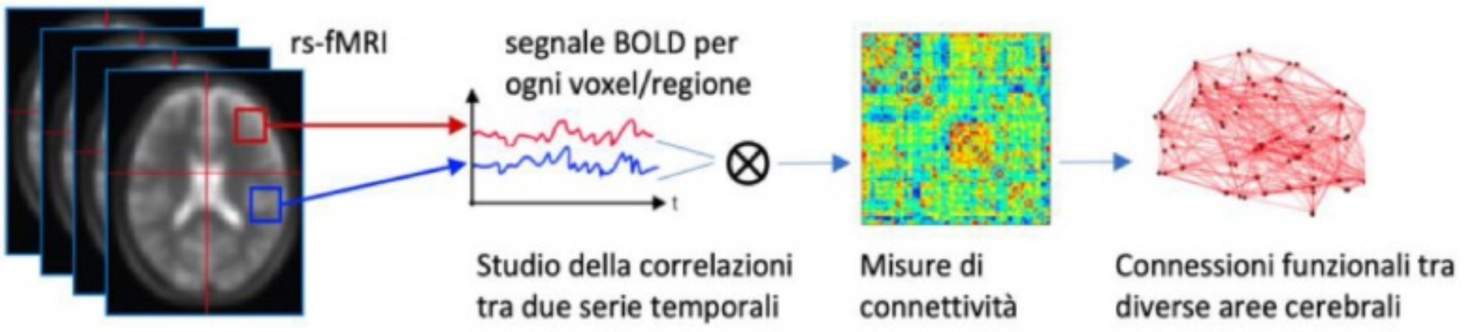
Analysis on ABIDE I preprocessed

CPAC - Configurable Pipeline for the Analysis of Connectomes

We selected the **CPAC** pipeline, since it's the most widely used in recent literature.

CPAC is a configurable open-source software, based on Nipype project that gather tools such as AFNI, ANTS, FLS for rs -fMRI image processing, like time slicing, motion correction, normalization and registration.

Data are standardized to MNI space and temporal series are extracted using different **atlases** like AAL, HO, CC-200, CC-400



Atlases

We investigated several atlases:

- Harvard-Oxford (HO): 110 ROIs
- Automated Anatomical Labeling (AAL): 116 ROIs
- Craddock 200 (CC200): 200 ROIs
- Craddock 400 (CC400): 392 ROIs

From each ROI the averaged temporal series is exctracted.

Connectivity metrics

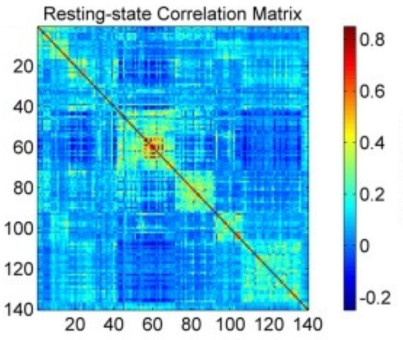
On preprocessed data, functional connectivity is evaluated by means of different metrics:

○ Pearson correlation:

$$r_{xy} = \frac{\sum_{i=1}^{N}(x_i - \hat{x})(y_i - \hat{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \hat{x})^2 \sum_{i=1}^{N}(y_i - \hat{y})^2}}$$

Fisher transform:

$$Z_{xy} = \sqrt{N-3}\,\tanh^{-1}(r_{xy})$$



○ Partial correlation (first order):

$$r_{xy.z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1-r_{xz}^2)(1-r_{yz}^2)}}$$

○ Tangent: covariance estimates are projected onto a common tangent space to reduce the statistical dependencies between their elements.

○ Precision: the inverse of the covariance matrix

All metrics are calculated using the nilearn[1] python package

---

[1]) *Nilearn: Statistics for NeuroImaging in Python* - https://nilearn.github.io

The curse of dimensionality

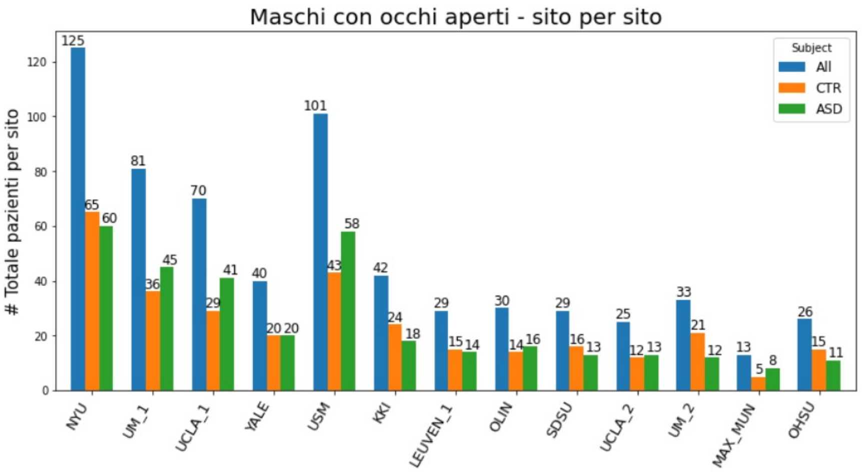
Best case scenario:

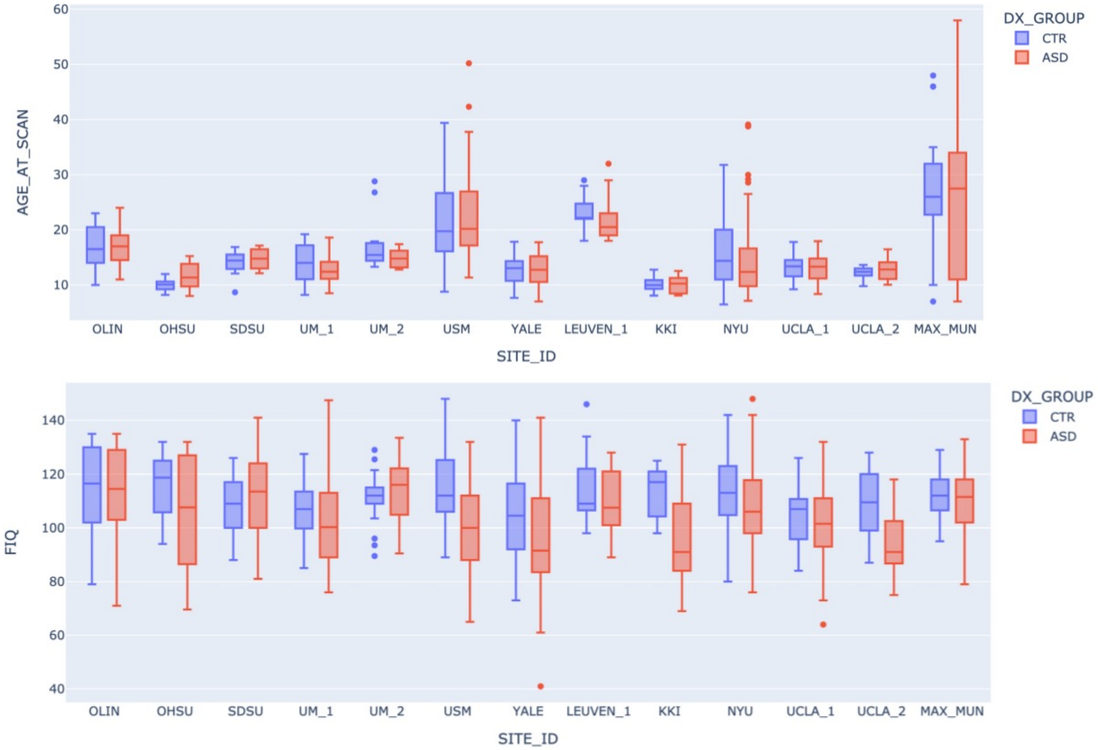
- HO atlas: 110 ROIs => Connetivity matrix ~ 6000 features
- ABIDE I: 1000 subjects…..

Single-site study:

- NYU: 184 subjects…

Dataset heterogeneity





Dataset heterogeneity

Inclusion criteria:

- Males
- Eyes open
- On *Mean Framewise Displacement* , $|X_i - \text{median}(X)| < 3 \cdot \text{MAD}$

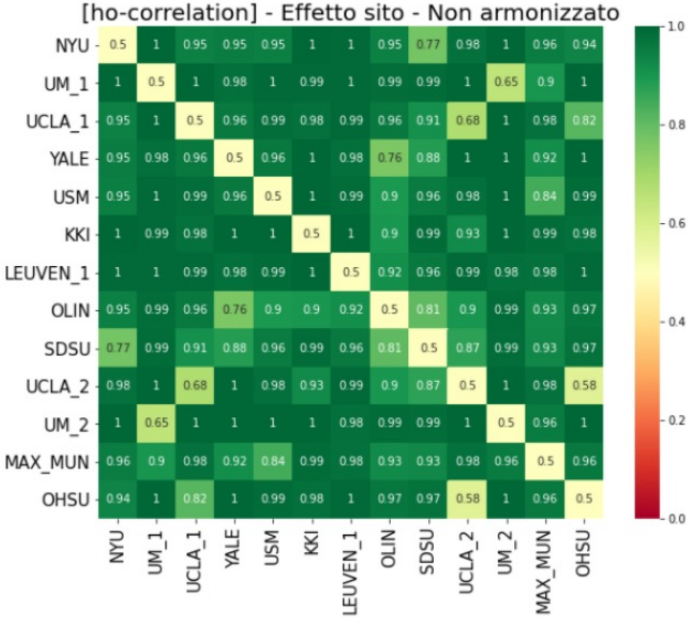
Resulting dataset:

580 subjects, 284 ASD and 196 controls.

Dataset heterogeneity

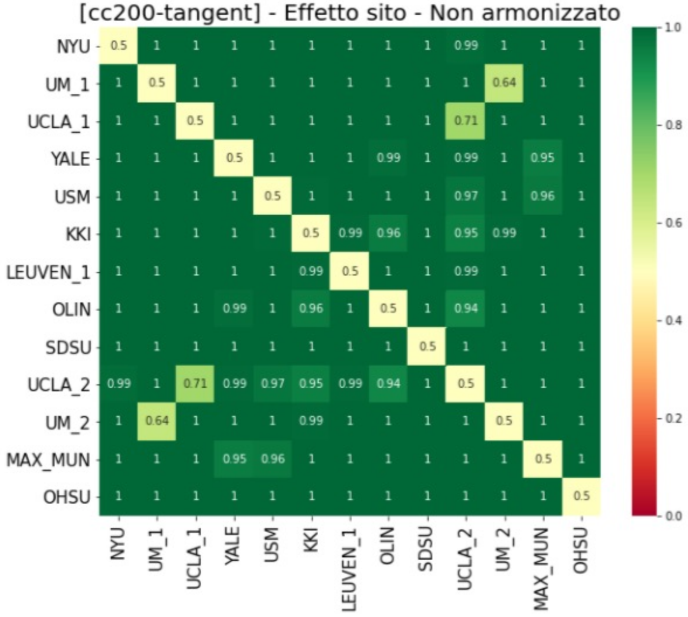
Site effect:

SVM rbf, trained to classify Site 1 vs Site 2 from connectivity features (only controls). AUC scores.



[ho-correlation] - Effetto sito - Non armonizzato

	NYU	UM_1	UCLA_1	YALE	USM	KKI	LEUVEN_1	OLIN	SDSU	UCLA_2	UM_2	MAX_MUN	OHSU
NYU	0.5	1	0.95	0.95	0.95	1	1	0.95	0.77	0.98	1	0.96	0.94
UM_1	1	0.5	1	0.98	1	0.99	1	0.99	0.99	1	0.65	0.9	1
UCLA_1	0.95	1	0.5	0.96	0.99	0.98	0.99	0.96	0.91	0.68	1	0.98	0.82
YALE	0.95	0.98	0.96	0.5	0.96	1	0.98	0.76	0.88	1	1	0.92	1
USM	0.95	1	0.99	0.96	0.5	1	0.99	0.9	0.96	0.98	1	0.84	0.99
KKI	1	0.99	0.98	1	1	0.5	1	0.9	0.99	0.93	1	0.99	0.98
LEUVEN_1	1	1	0.99	0.98	0.99	1	0.5	0.92	0.96	0.99	0.98	0.98	1
OLIN	0.95	0.99	0.96	0.76	0.9	0.9	0.92	0.5	0.81	0.9	0.99	0.93	0.97
SDSU	0.77	0.99	0.91	0.88	0.96	0.99	0.96	0.81	0.5	0.87	0.99	0.93	0.97
UCLA_2	0.98	1	0.68	1	0.98	0.93	0.99	0.9	0.87	0.5	1	0.98	0.58
UM_2	1	0.65	1	1	1	1	0.98	0.99	0.99	1	0.5	0.96	1
MAX_MUN	0.96	0.9	0.98	0.92	0.84	0.99	0.98	0.93	0.93	0.98	0.96	0.5	0.96
OHSU	0.94	1	0.82	1	0.99	0.98	1	0.97	0.97	0.58	1	0.96	0.5



[cc200-tangent] - Effetto sito - Non armonizzato

	NYU	UM_1	UCLA_1	YALE	USM	KKI	LEUVEN_1	OLIN	SDSU	UCLA_2	UM_2	MAX_MUN	OHSU
NYU	0.5	1	1	1	1	1	1	1	1	0.99	1	1	1
UM_1	1	0.5	1	1	1	1	1	1	1	1	0.64	1	1
UCLA_1	1	1	0.5	1	1	1	1	1	1	0.71	1	1	1
YALE	1	1	1	0.5	1	1	1	0.99	1	0.99	1	0.95	1
USM	1	1	1	1	0.5	1	1	1	1	0.97	1	0.96	1
KKI	1	1	1	1	1	0.5	0.99	0.96	1	0.95	0.99	1	1
LEUVEN_1	1	1	1	1	1	0.99	0.5	1	1	0.99	1	1	1
OLIN	1	1	1	0.99	1	0.96	1	0.5	1	0.94	1	1	1
SDSU	1	1	1	1	1	1	1	1	0.5	1	1	1	1
UCLA_2	0.99	1	0.71	0.99	0.97	0.95	0.99	0.94	1	0.5	1	1	1
UM_2	1	0.64	1	1	1	0.99	1	1	1	1	0.5	1	1
MAX_MUN	1	1	1	0.95	0.96	1	1	1	1	1	1	0.5	1
OHSU	1	1	1	1	1	1	1	1	1	1	1	1	0.5

Harmonization: NeuroHarmonize

These data suffer from technical between-scanner (namely between-site) variations, they hence need to be harmonized.

A common approach in dealing with site dependence employs **ComBat-GAM**: a batch-effect correction tool aimed to reduce inter-site variability while preserving variations due to other biologically-relevant covariates[1].

For site *i*, subject *j*, region *k*:

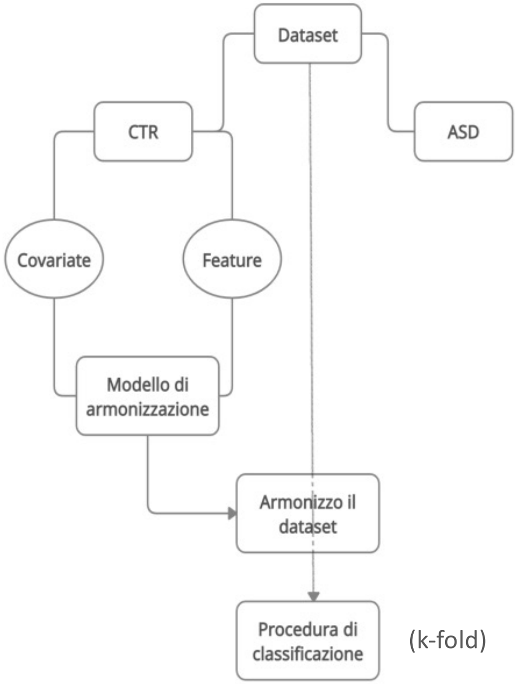
$$Y^*_{ijk} = (Y_{ijk} - f_k\,(x_{ij}, z_{ij}, w_{ij}) - g^*_{ik}) \,/\, d^*_{ik} + f_k\,(x_{ij}, z_{ij}, w_{ij})$$

---

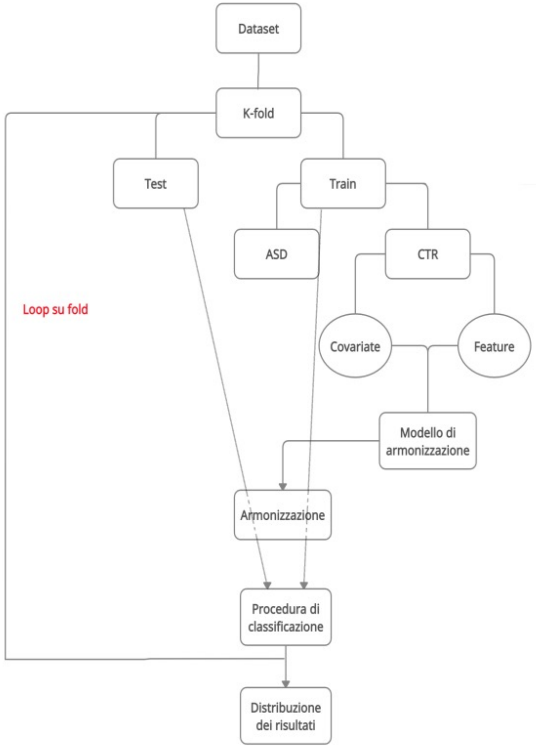
[1]) Pomponio, R., et al., (2020). "Harmonization of large MRI datasets for the analysis of brain imaging patterns throughout the lifespan", Neuroimage 208, 116450

Harmonization strategies

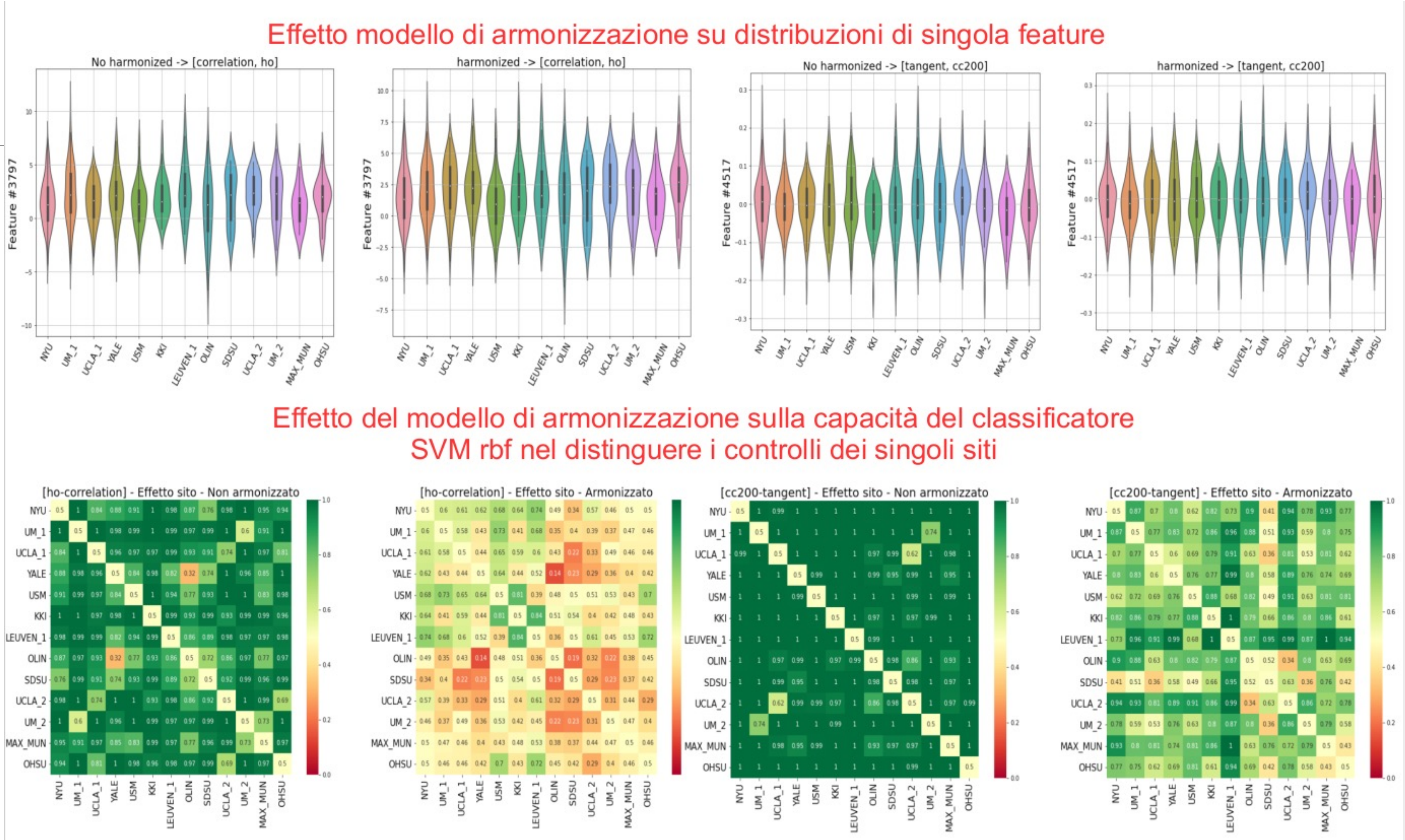
All the controls are used to harmonize


Dataset
CTR
ASD
Covariate
Feature
Modello di armonizzazione
Armonizzo il dataset
Procedura di classificazione
(k-fold)

Only the controls in the training set are used to harmonize


Dataset
K-fold
Test
Train
ASD
CTR
Loop su fold
Covariate
Feature
Modello di armonizzazione
Armonizzazione
Procedura di classificazione
Distribuzione dei risultati

Harmonization using all the controls

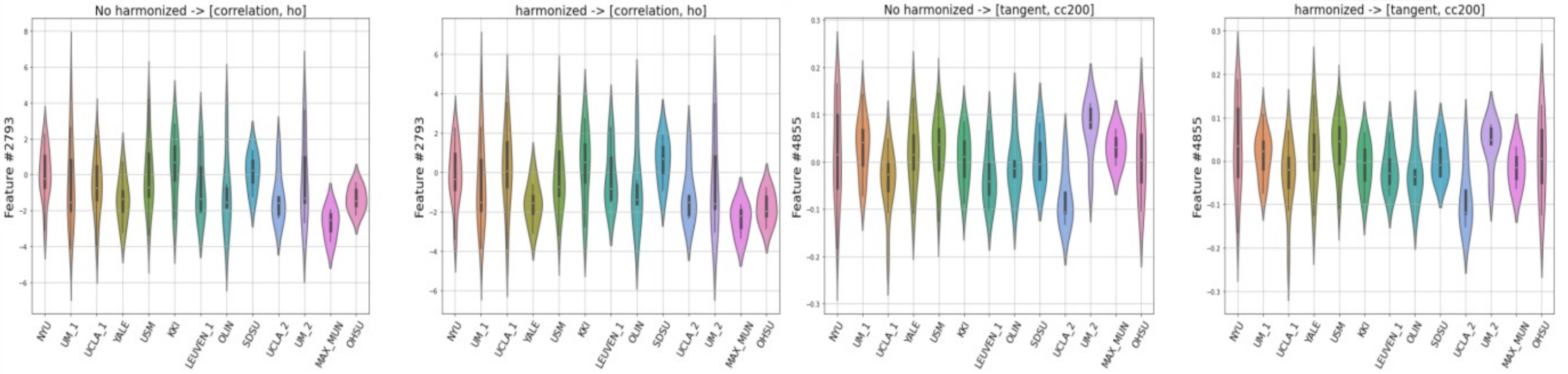


Effetto modello di armonizzazione su distribuzioni di singola feature

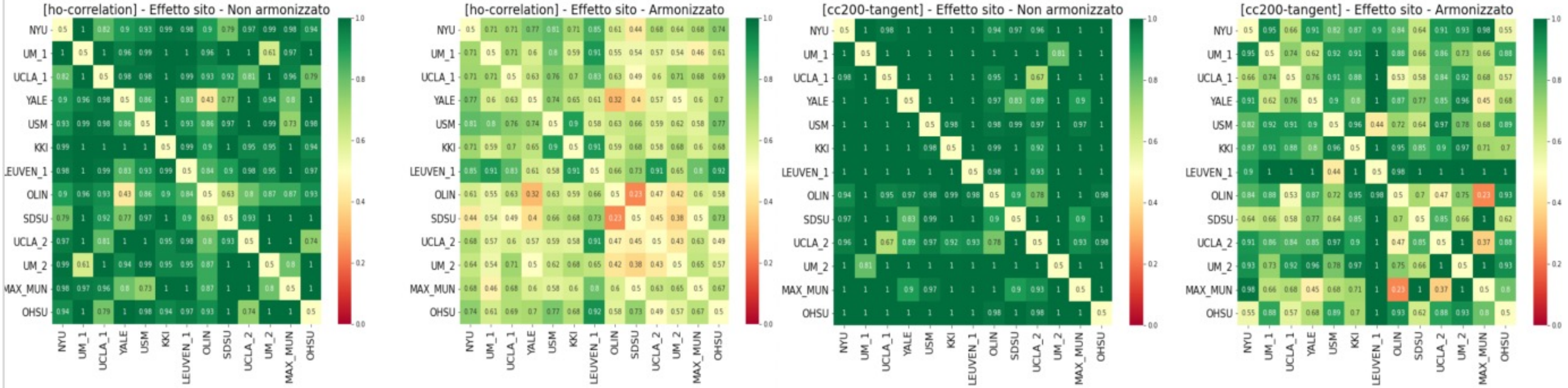
Effetto del modello di armonizzazione sulla capacità del classificatore SVM rbf nel distinguere i controlli dei singoli siti

Harmonization using only the controls in the training set

Effetto modello di armonizzazione su distribuzioni di singola feature

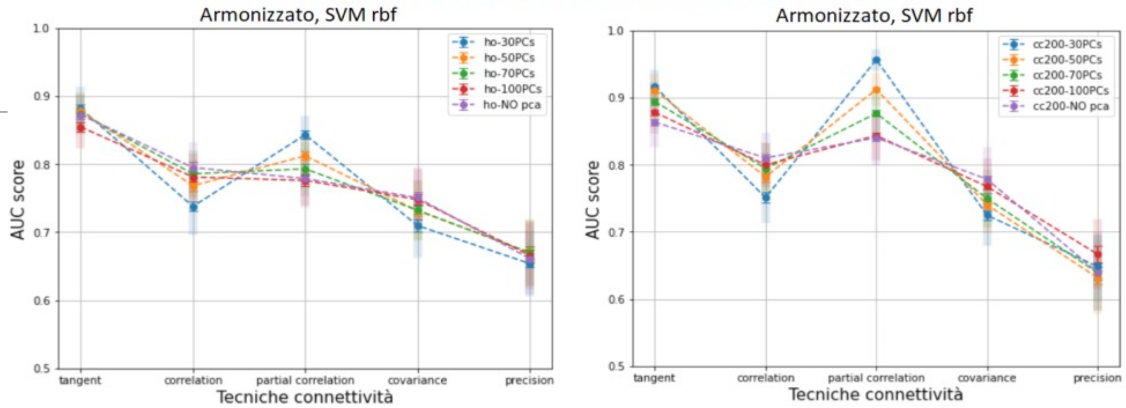


Effetto del modello di armonizzazione sulla capacità del classificatore
SVM rbf nel distinguere i controlli dei singoli siti

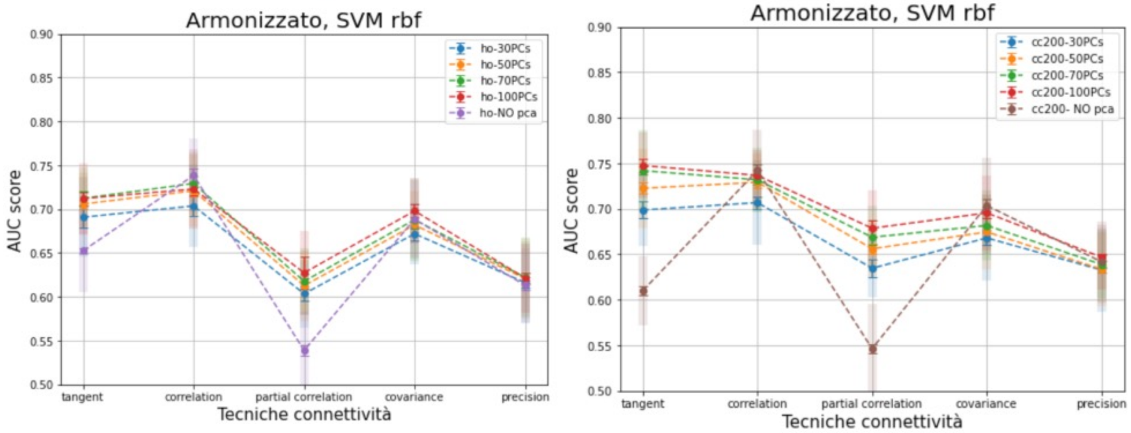


Harmonization strategies: results

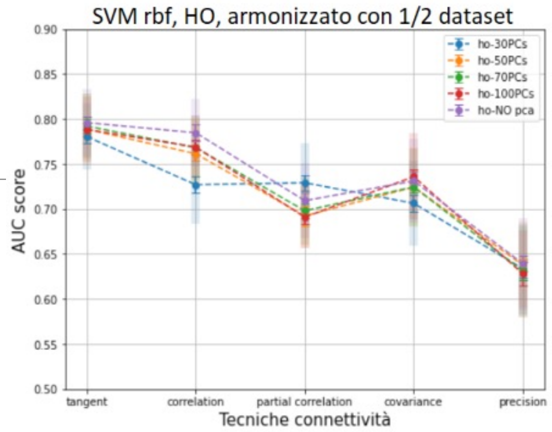
Stima parametri del modello con l'intero dataset

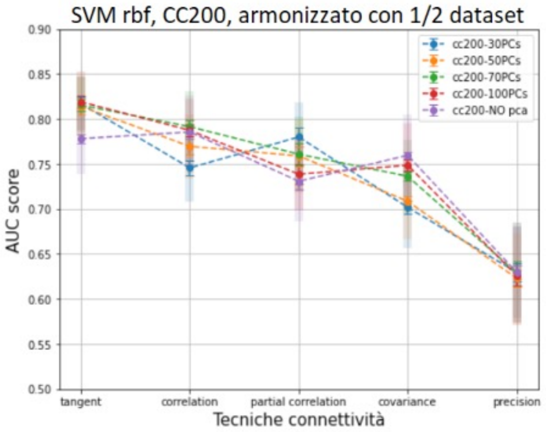


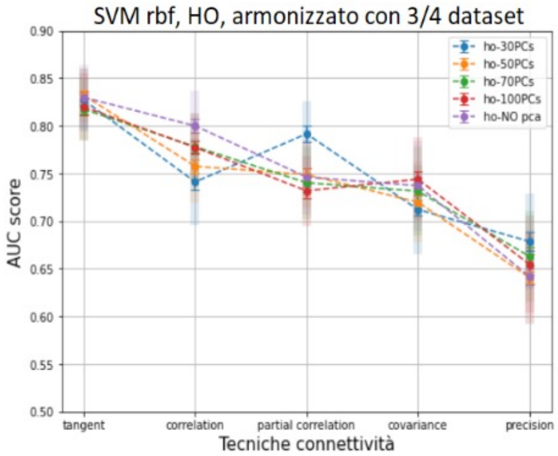
Stima parametri del modello con il training set

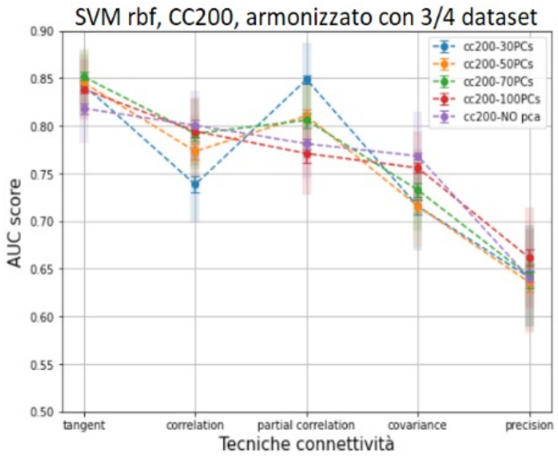


Bias effect? (2)

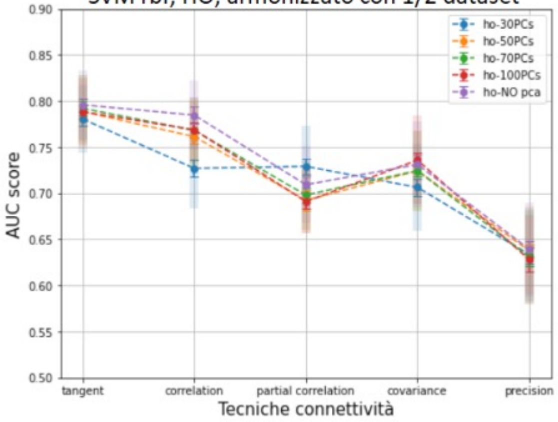




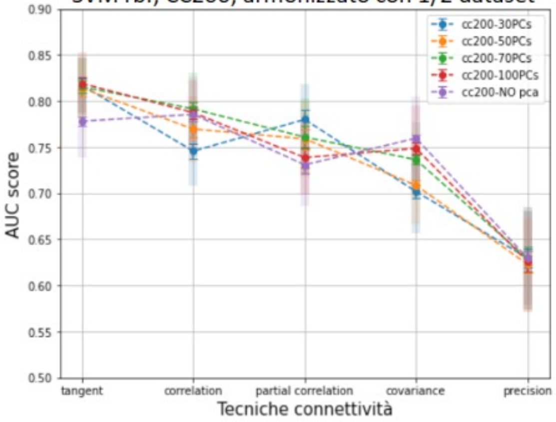




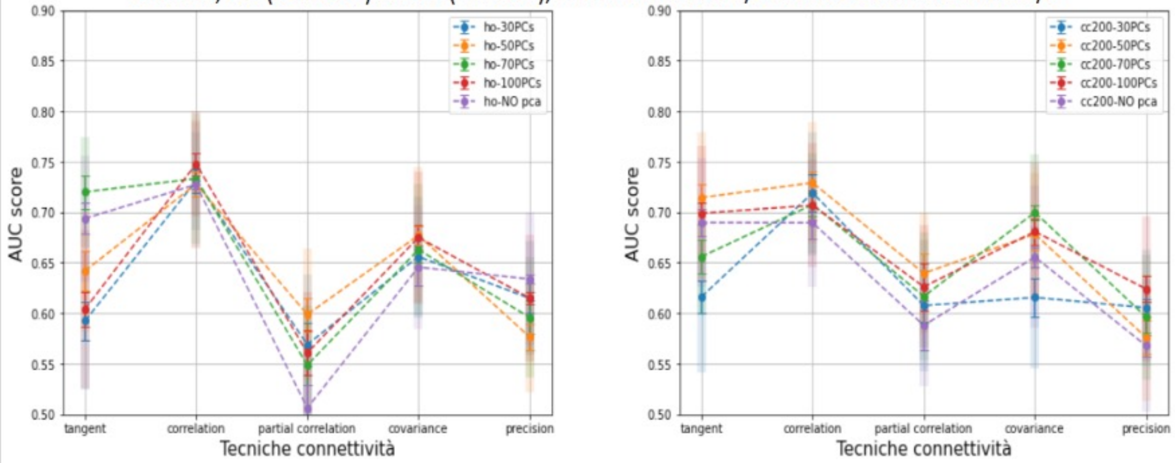
Bias effect?



SVM rbf, HO, armonizzato con 1/2 dataset

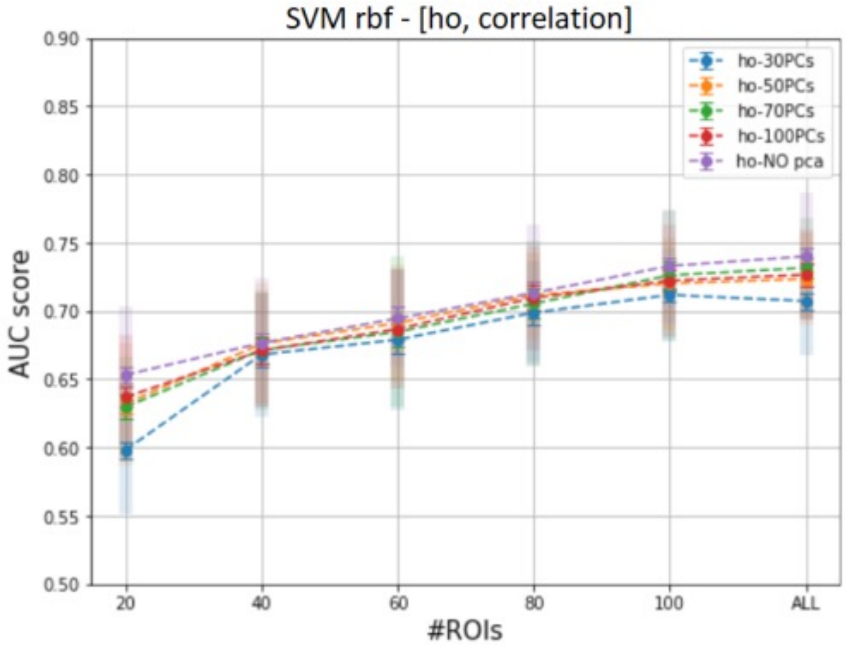


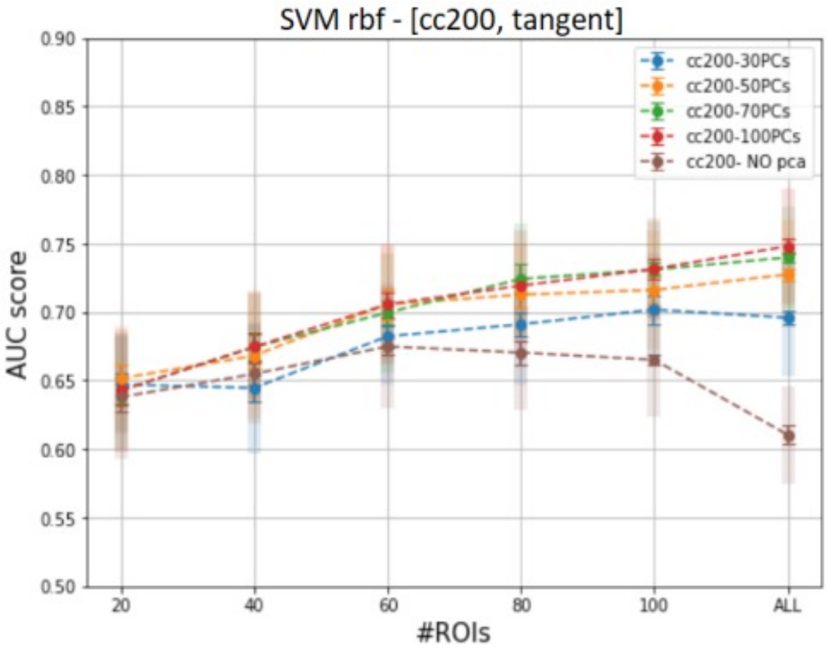
SVM rbf, CC200, armonizzato con 1/2 dataset



SVM rbf, HO (a sinistra) CC200 (a destra), armonizzato con 1/2 dataset e validato l'altro 1/2

Dependence on ROIs and PCs





ABIDE I & II – Data processing

ABIDE I & II – Data processing

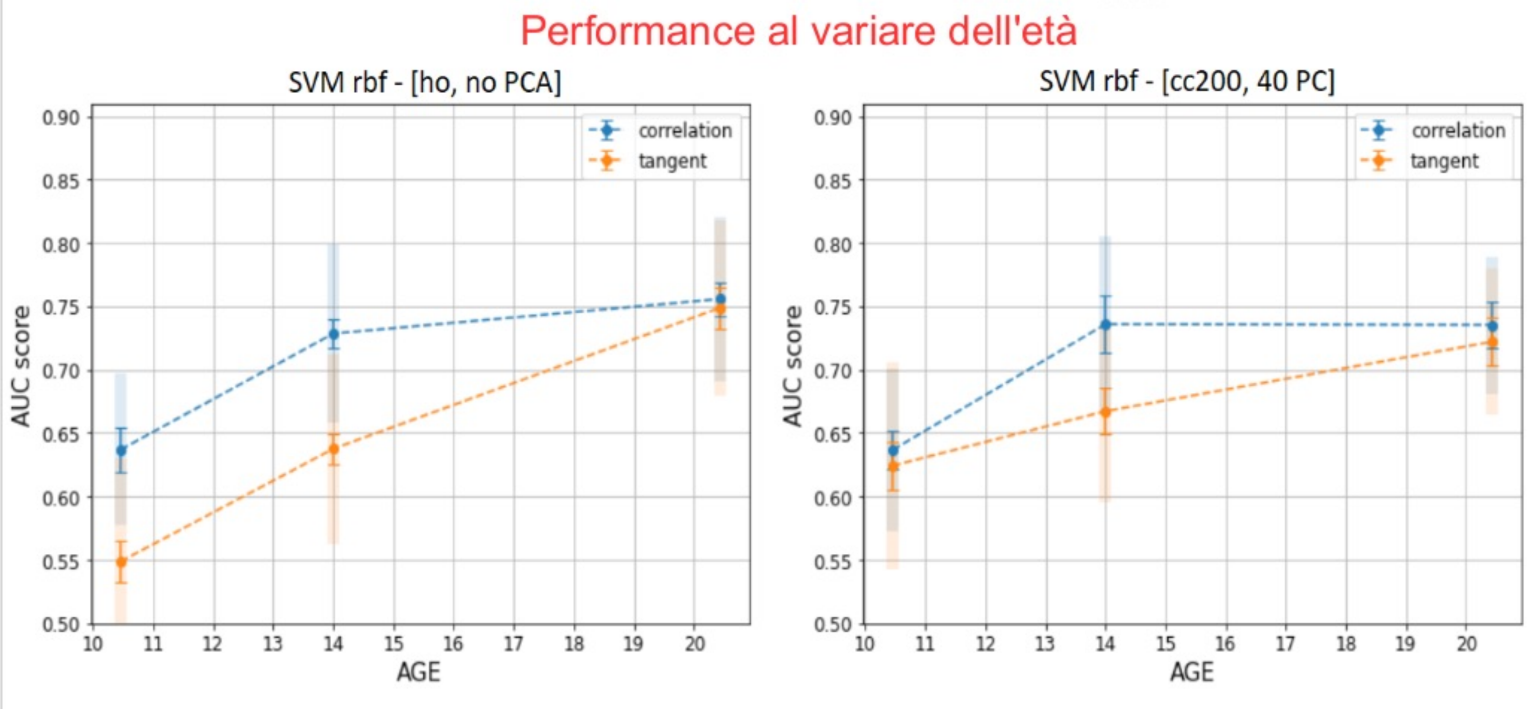
We are currently processing the whole ABIDE dataset (I&II) using the C-PAC pipeline

- ABIDE I: done

- ABIDE II: in progress (more than 570% done, to be finished in the next days)

The result will allow us will allow us to fully exploit the potentiality of the dataset, using homogeneusly processed data

Age dependence



Performance al variare dell'età

SVM rbf - [ho, no PCA]

SVM rbf - [cc200, 40 PC]

3 age bins (equally populated):

- age ≤ 12
- 12 < age ≤ 16.3
- age > 16.3

Future work

On ABIDE, using the full dataset:

Harmonization strategies

- NeuroHarmonize
- New strategies based on adversarial networks

Dimensionality reduction

- Different atlases
- Feature selection

Dinamic connectivity…

Effect of TMS in treatment of substance use disorders