Lightning Talk

Low Rank Mixup Augmentations for Contrastive Learning of Phenotypes from Functional Connectivity

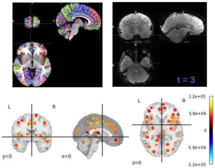
Anton Orlichenko, Gang Qu, Ziyu Zhou, Yu-Ping Wang

[1]Department of Biomedical Engineering, Tulane University, New Orleans, LA

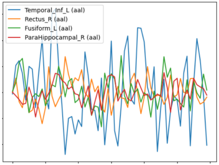
MCBIOS 2024, Emory University, March 22-24

TULANE UNIVERSITY

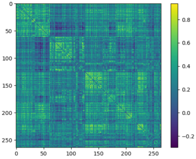
fMRI Machine Learning Pipeline



fMRI and Atlas



BOLD Timeseries



Functional Connectivity

Phenotypes

1. Age
2. Sex
3. Race
4. Schizophrenia
5. Neurodegen Disease

- How can we improve the accuracy of phenotype predictions?
- **Try contrastive learning**
  - Maximize similarity between **positive pairs**
  - Minimize similarity between **negative pairs**

TULANE UNIVERSITY

Motivation

Problem

Most contrastive learning frameworks require large numbers of subjects or data augmentations

- Most fMRI studies recruit fewer than 100 subjects
- Not clear how to augment fMRI-derived metrics such as functional connectivity (FC)

Solution

We create an augmentation strategy for FC based on mixup of the rank-1 approximation of FC (first component)

- First component is not effective for phenotype prediction

TULANE UNIVERSITY

Idea

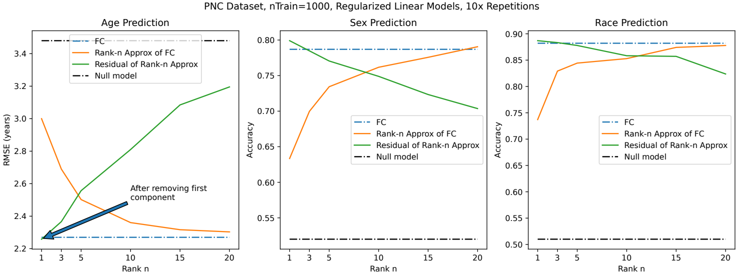
- Construct low rank approximation to FC $\mathbf{X}$ using the eigendecomposition
- Only keep the top N eigenvalues; set the rest to zero
- Since FC is a symmetric (PSD) matrix, it is orthogonally diagonalizable

$$\begin{aligned} \mathbf{X} &= \mathbf{V\Lambda V}^{\mathsf{T}} \\ \hat{\lambda}_{ii} &= \begin{cases} \lambda_{ii}, & i \leq N \\ 0, & \text{otherwise} \end{cases} \\ \mathbf{X}^{(N)} &= \mathbf{V\hat{\Lambda} V}^{\mathsf{T}} \end{aligned} \tag{1}$$

TULANE UNIVERSITY

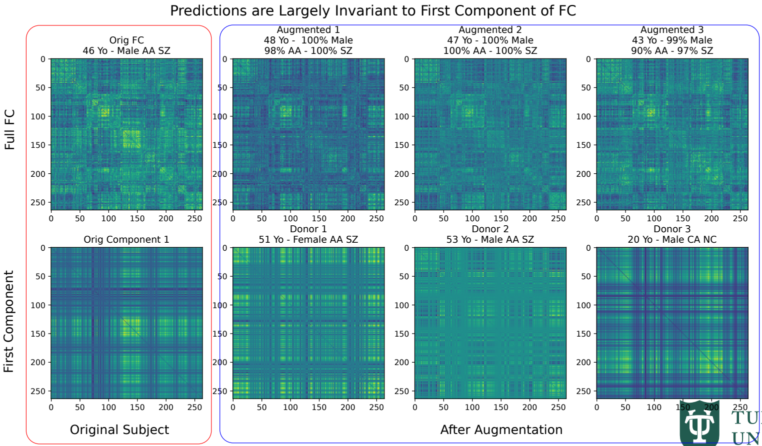
Removing the First Component

- Removing the first component doesn't reduce prediction accuracy (at all)
- Green curve at left in all graphs



Mixup Augmentations

- Use the ineffective first component in mixup augmentations to create positive pairs



TULANE UNIVERSITY

Moderately Improved Prediction Results

- Prediction results are improved 2-10% over MLP and GCN models not using augmentations
- Using PNC and BSNIP datasets

Dataset	Phenotype	Metric	MLP	GCN	CL+LRAug	p-value
BSNIP	Age	RMSE (yr)	11.82 ± 0.62	11.07 ± 0.67	$\mathbf{10.25 \pm 0.64}$	<0.001
BSNIP	Sex	Accuracy	68.1 ± 4.6	66.4 ± 6.1	$\mathbf{71.7 \pm 5.1}$	0.003
BSNIP	Race	Accuracy	76.0 ± 3.8	72.4 ± 6.1	$\mathbf{77.7 \pm 4.4}$	0.026
BSNIP	SZ	Accuracy	75.2 ± 3.6	71.2 ± 6.7	$\mathbf{76.9 \pm 4.4}$	0.017
PNC	Age	RMSE (yr)	2.62 ± 0.14	2.44 ± 0.12	$\mathbf{2.18 \pm 0.07}$	<0.001
PNC	Sex	Accuracy	77.9 ± 2.0	73.7 ± 9.8	$\mathbf{79.7 \pm 2.2}$	<0.001
PNC	Race	Accuracy	87.7 ± 1.6	86.6 ± 3.8	$\mathbf{89.8 \pm 1.9}$	<0.001

TULANE UNIVERSITY

- Removing the first component of FC helps identify same subject from different scan better than raw FC
- 97.3% identification accuracy versus 62.5%