A Denoising Variational Autoencoder for the diagnosis of Autism based on Resting-state fMRI

Columbia University Data Science Institute

Xinyuan Zheng[3,4], Orren Ravid[2], Yoojean Kim[2], Xi Zhu[1,2] Xiaofu He[1,2,4*]
[1]Department of Psychiatry, [2]The New York Psychiatric Institute, [3] Department of Statistics,
[4]Data Science Institute, Columbia University

Introduction

- Autism spectrum disorders (ASD) are neurodevelopmental disorders affect the perception, social behavior and communication of the patients. As ASD has complexity of phenotypes, it is a challenge to find clinically useful diagnostic biomarkers and objectively diagnose patients with ASD.

- In this project, we explored classification of the resting-state fMRI (rs-fMRI) for ASD using a Denoising Variational AutoEncoder (DVAE).

Data

- We used the dataset (publicly available at: https://paris-saclay-cds.github.io/autism_challenge/) from Paris-Saclay Center for Data Science that was initially published for competition in the IMaging-PsychiAtry Challenge (IMPAC).

- This dataset contained 601 health controls and 549 ASD patients with age range from 5 to 64 and a summary of the demographics is shown in Figure 1.

- The dataset is split into training and testing (80:20).

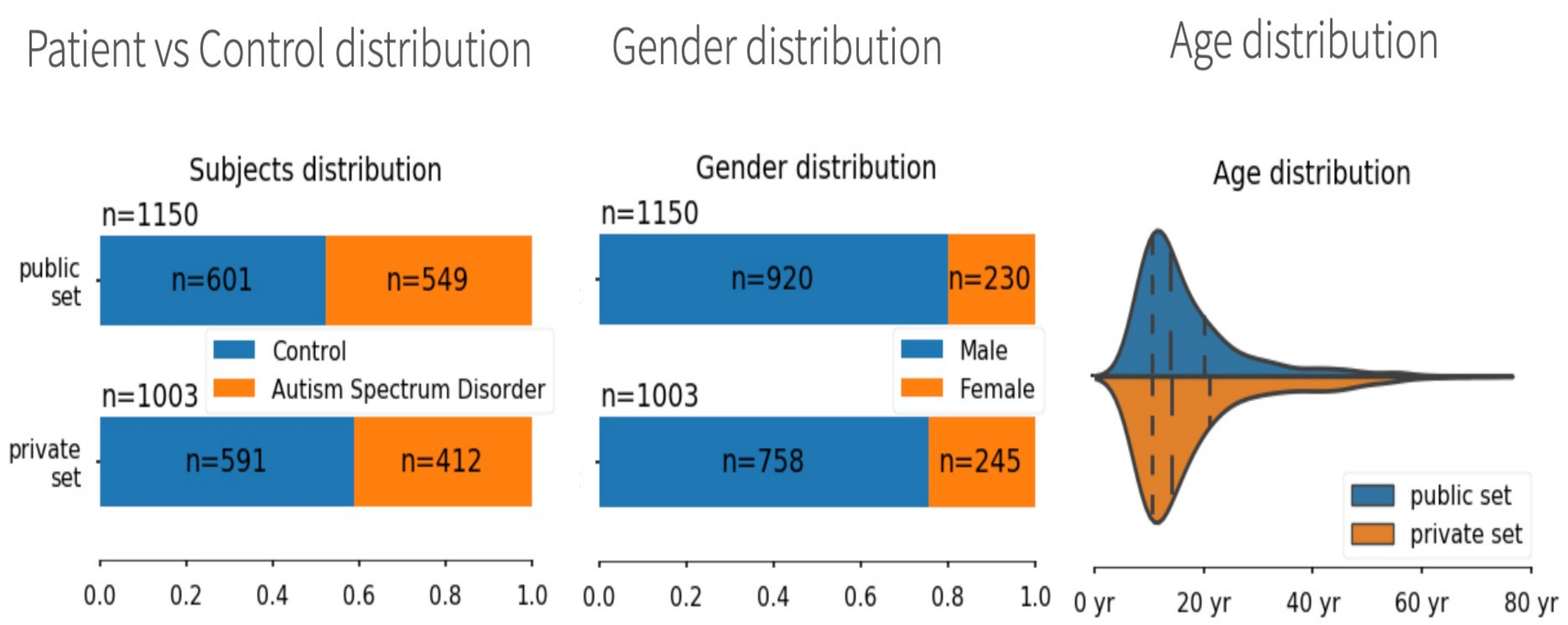

Figure 1. Distribution of the Autism Dataset. (R. Toro et al. 2018).

Method: DVAE + ML classifiers

Feature Extraction:

- Functional connectivity matrix (249x249) is firstly extracted from rs-fMRI using Ncuts parcellations with 249 ROIs (Craddock 2012).
- A fine-tuned DVAE is used to reduce the dimensionality and extract 10 features (5 latent distributions) from the flattened connectivity vector. This procedure is shown in Figure 2.

Classifier:

- ML classifiers including SVM and Random Forest are used to classify the extracted latent variables into ASD patient and healthy control.
- Grid search and 10-fold cross validation are used to find best parameters for the classifiers.
- Thresholds of the classifiers were tuned by maximizing the geometric mean of the sensitivity and specificity.

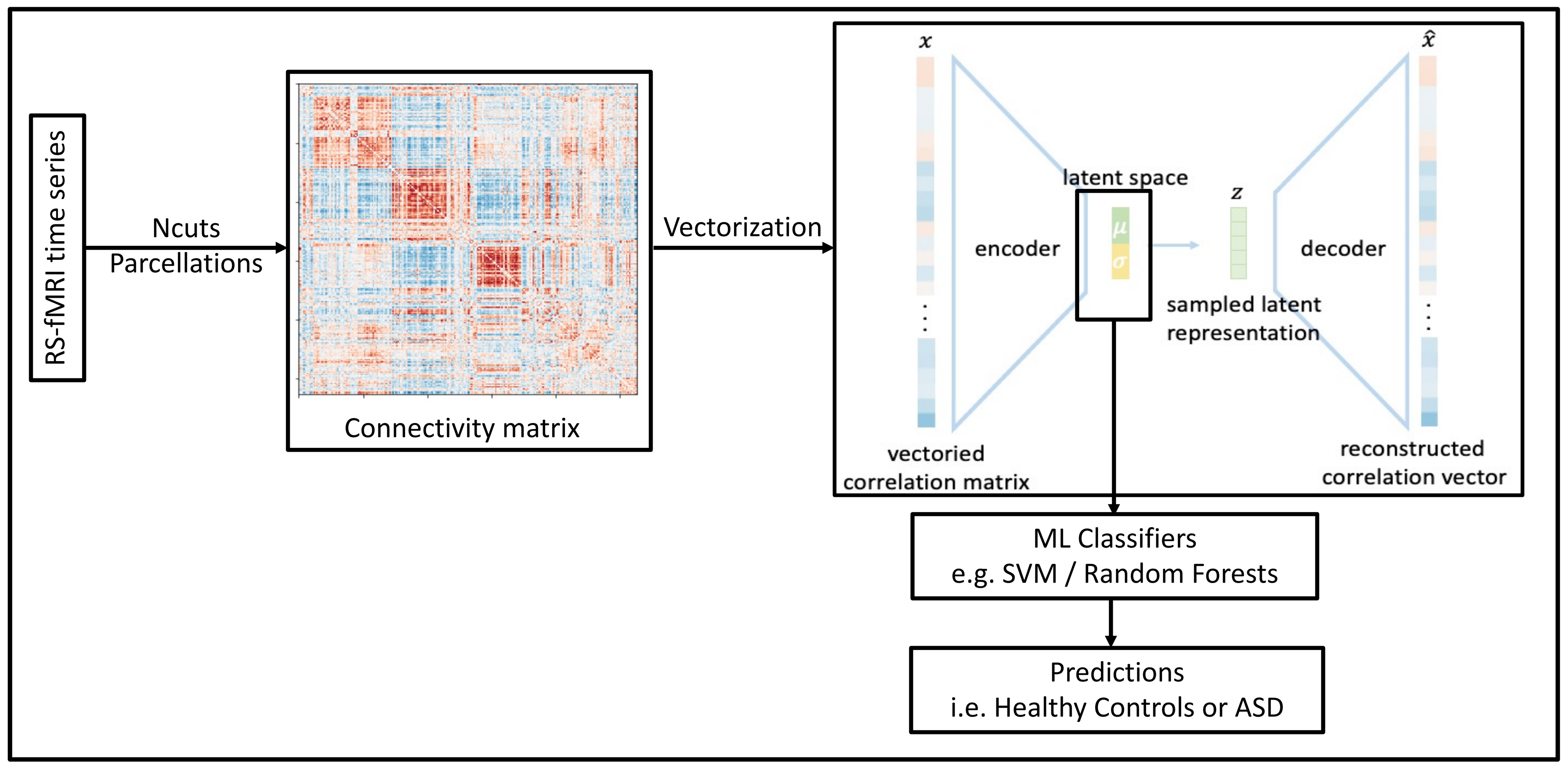

Figure 2. Classification Pipeline using RS-fMRI for ASD Diagnosis

		DVAE + ML clf	ML clf
Random Forest	train accuracy	0.62	0.86
	test accuracy	0.59	0.57
	test sensitivity	0.55	0.51
	test specificity	0.63	0.62
	test AUC	0.57	0.56
SVM	train accuracy	0.57	0.94
	test accuracy	0.58	0.6
	test sensitivity	0.62	0.73
	test specificity	0.57	0.49
	test AUC	0.60	0.65

Table 1. Classification Performance With and Without Feature Reduction

Result and Discussion

- The resulting performance are shown in Table 1.
- The training (blue) and validation (orange) loss of the DVAE is shown in Figure 3.
- The latent variables extracted from the DVAE achieved comparable results with far less features.
- The proposed model efficiently alleviated the overfitting problem exists in traditional ML classifiers, reduced the dimensionality of the featuers, accelerated training for the classifiers and testing for new subjects, and made assessing the confidence intervals of the classification results possible.
- Future work could potentially rectify the low accuracy by training separate models for controls and patients and taking the differences across sites into account, visualization of the latents and the model outcomes could also help tune the hyperparameters better and locate the most different brain regions hence improve the interpretability.

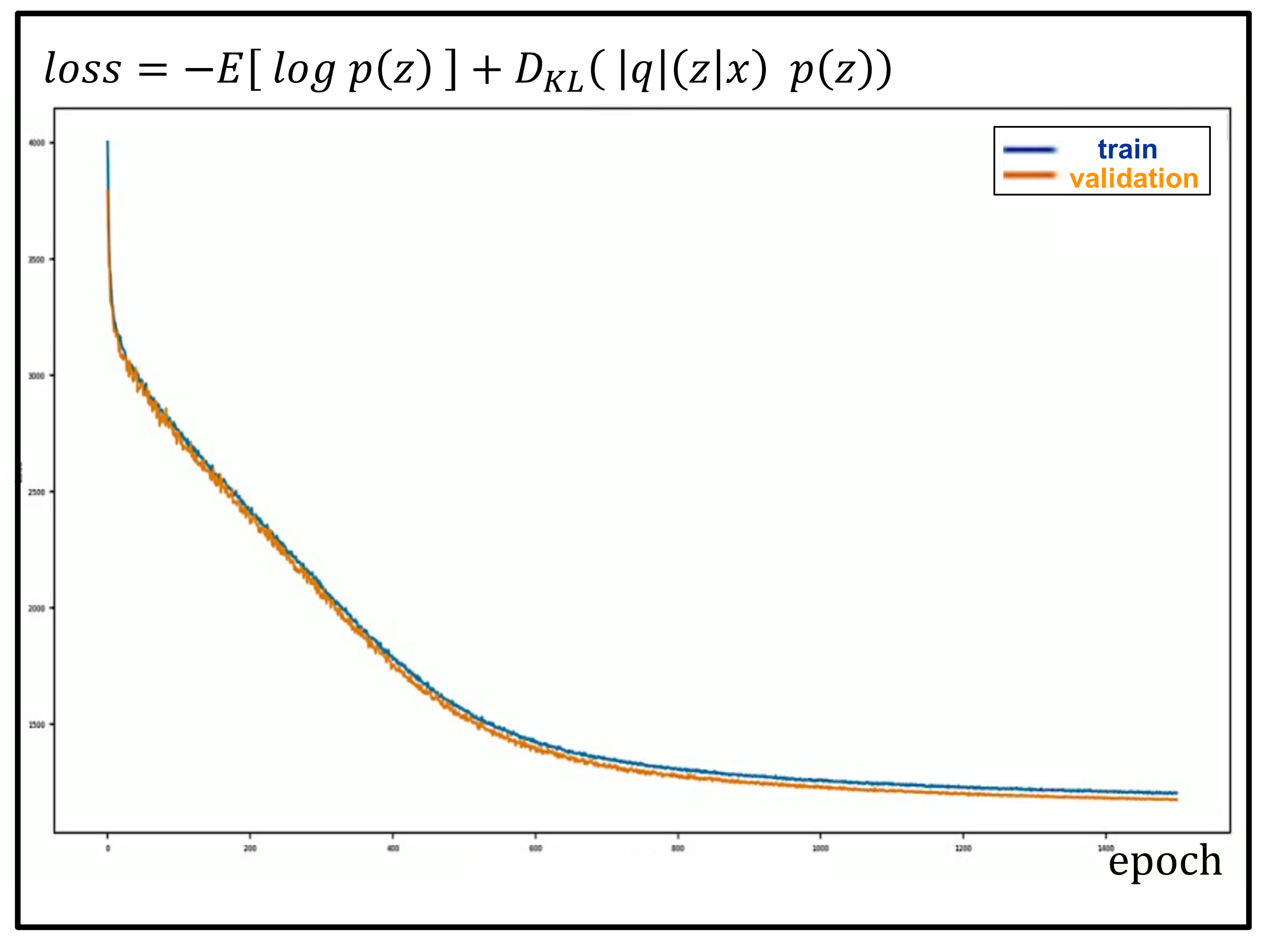

Figure 3. DVAE Training (blue) and Validation (orange) Loss vs. Epoch

References

Roberto Toro et al. 2018. “Imaging-psychiatry challenge: predicting autism, A data challenge on Autism Spectrum Disorder detection.” https://paris-saclay-cds.github.io/autism_challenge/

Craddock, R Cameron et al. 2012. “A Whole Brain FMRI Atlas Generated via Spatially Constrained Spectral Clustering.” Human brain mapping 33(8): 1914–28.