

PIBE: Progression Invariant Brain Embeddings

Fermín Travi^{1,2}, Eduardo Castro³, Hongyang Li³, Anushree Mehta³, Jenna Reinen³, Amit Dhurandhar³, Pablo Meyer Rojas³, Guillermo A. Cecchi³, Pablo Polosecki³

1. Facultad de Ciencias Exactas y Naturales,
Departamento de Ciencias de la Computación,
Universidad de Buenos Aires, Buenos Aires, Argentina

2. Laboratorio de Inteligencia Artificial Aplicada (LIAA), CONICET
- Universidad de Buenos Aires, Instituto en Ciencias de la
Computación (ICC), Buenos Aires, Argentina

3. IBM T. J. Watson Research Center,
Yorktown Heights, New York, NY, United States
Computación (ICC), Buenos Aires, Argentina



ftravi@dc.uba.ar

Motivation

- One of the main hypotheses across a variety of disorders, such as **dementias** and schizophrenia, is that their effects produce **premature brain aging**.^[1]
- Finding correlations between brain markers and subjects' age **conflate phenotypes with progression** of the disorder.
- In this work, we leverage machine learning techniques to build **abstract brain representations (embeddings)** that are **invariant with respect to age**, thus avoiding the progression-phenotype confound by design.
- Were the premature brain aging hypothesis to be true, these embeddings should better predict the neurodegenerative disorders irrespective of the subjects' age.

Models

We build three different representations (z) of the brain, reconstructing brain **T1-weighted MRI scans** with a VAE^[3]:

- Age-invariant:** by reconstructing brain scans with a conditional decoder^[4], these embeddings are invariant with respect to age.
- Age-agnostic:** these embeddings are obtained without imposing any conditions on age.
- Age-variant:** contrary to the age-invariant representations, these are obtained by predicting age during the reconstruction.

All of them predicted the sex and BMI of the subject during training to enhance the retained information.

The encoder was based on a previously published 3D convolutional neural network employed for age prediction.

Datasets

General population

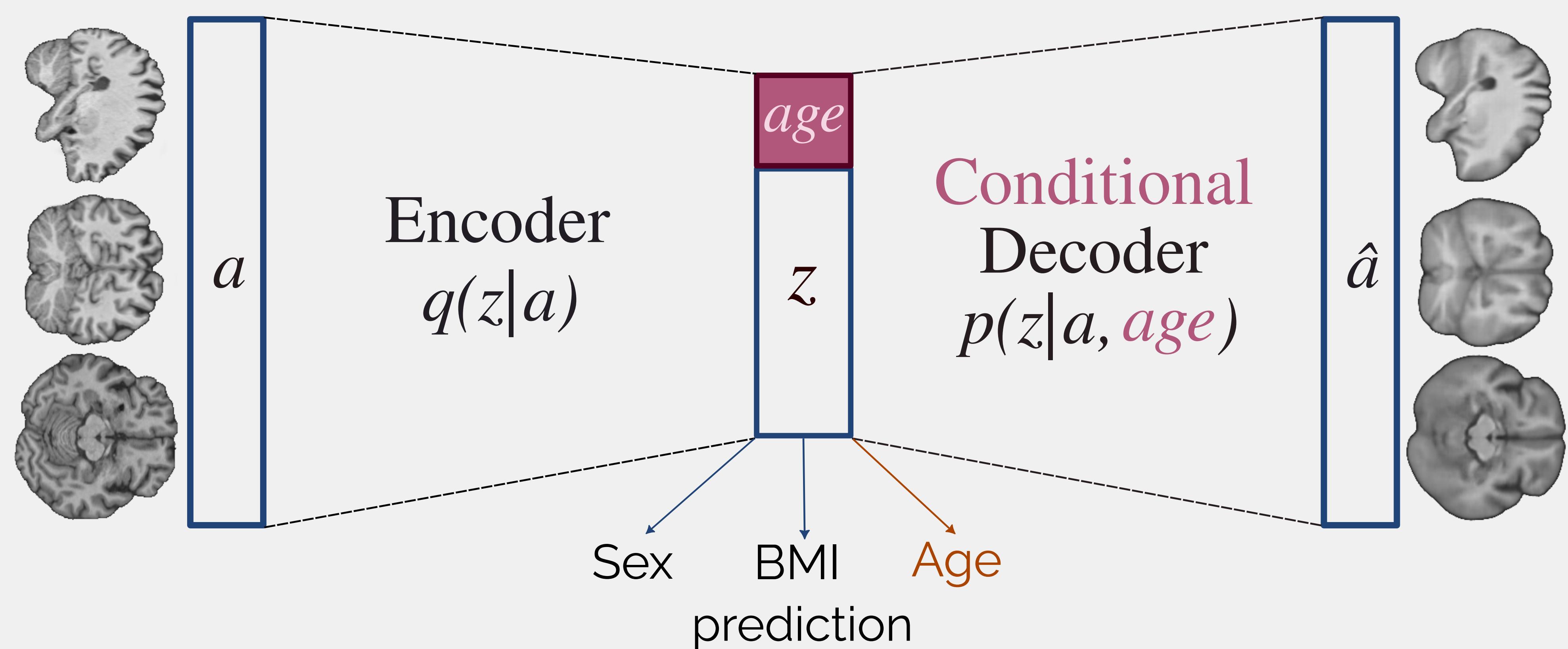
Split	Females	N	Mean age
Train	17,396	33,110	63 +/- 9.5
Validation	2,531	4,741	62 +/- 11.1
Test	3,495	6,631	63 +/- 9.2

- The **General population** dataset is a large scale dataset of T1-weighted (1mm³) brain MRI scans from the general population and is primarily used to build brain embedding models.
- It was obtained by bringing together, with a common preprocessing pipeline, brain scans from the UK-Biobank, CAM-CAN, SALD, NKIRS, HCP, and HCP-aging.^[2]

Diseased and healthy controls

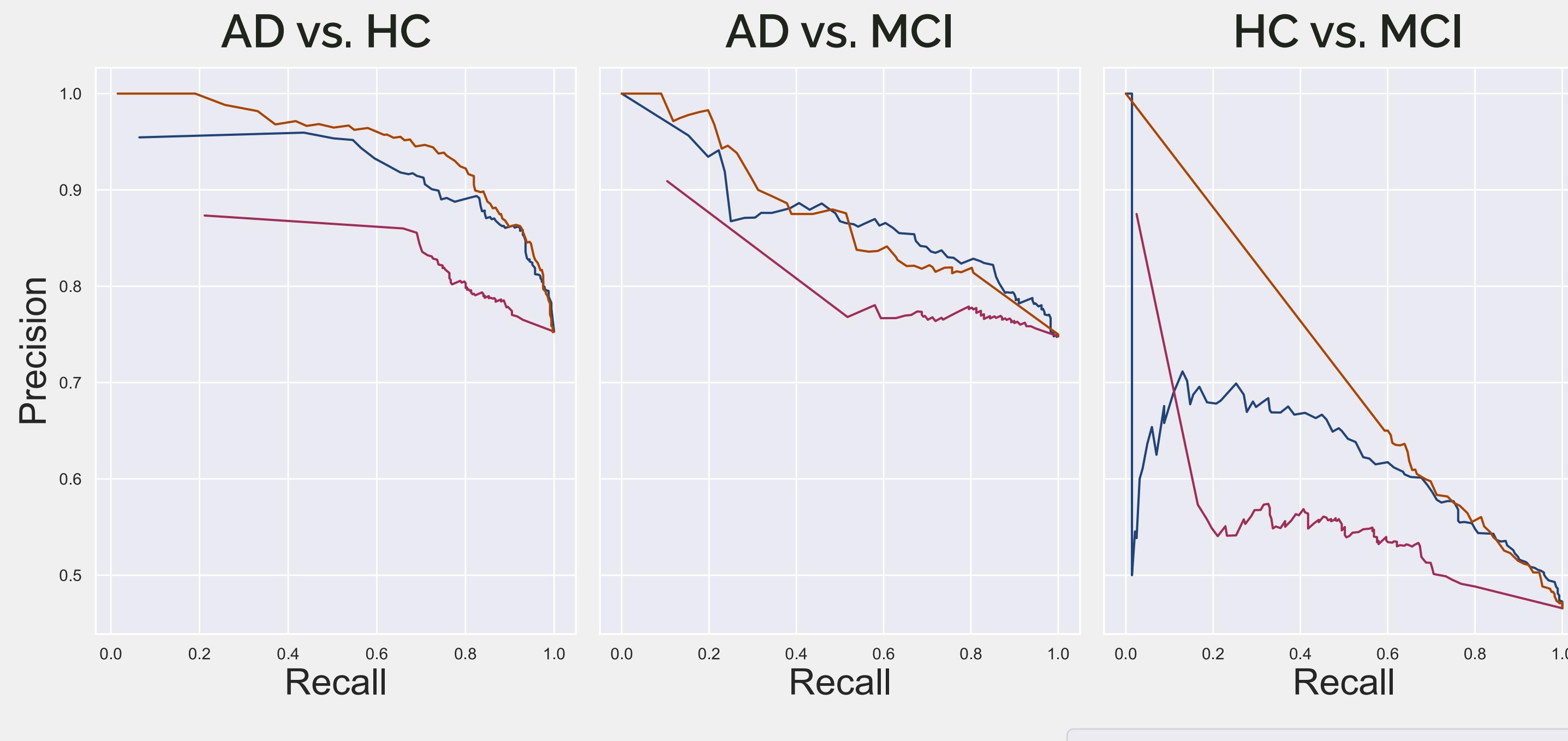
Group	Females	N	Mean age
AD	237	514	75.2 +/- 8.0
MCI	596	1430	73.1 +/- 7.5
HC	963	1713	70.5 +/- 8.0

- The **Diseased and healthy controls** dataset consists of T1-weighted (1mm³) brain MRI scans from individuals diagnosed with Mild Cognitive Impairment (MCI) or Alzheimer's Disease (AD) and their matched controls (HC).
- Following the same preprocessing steps as the General population dataset, it brings together brain scans from ADNI and OASIS.^[2]

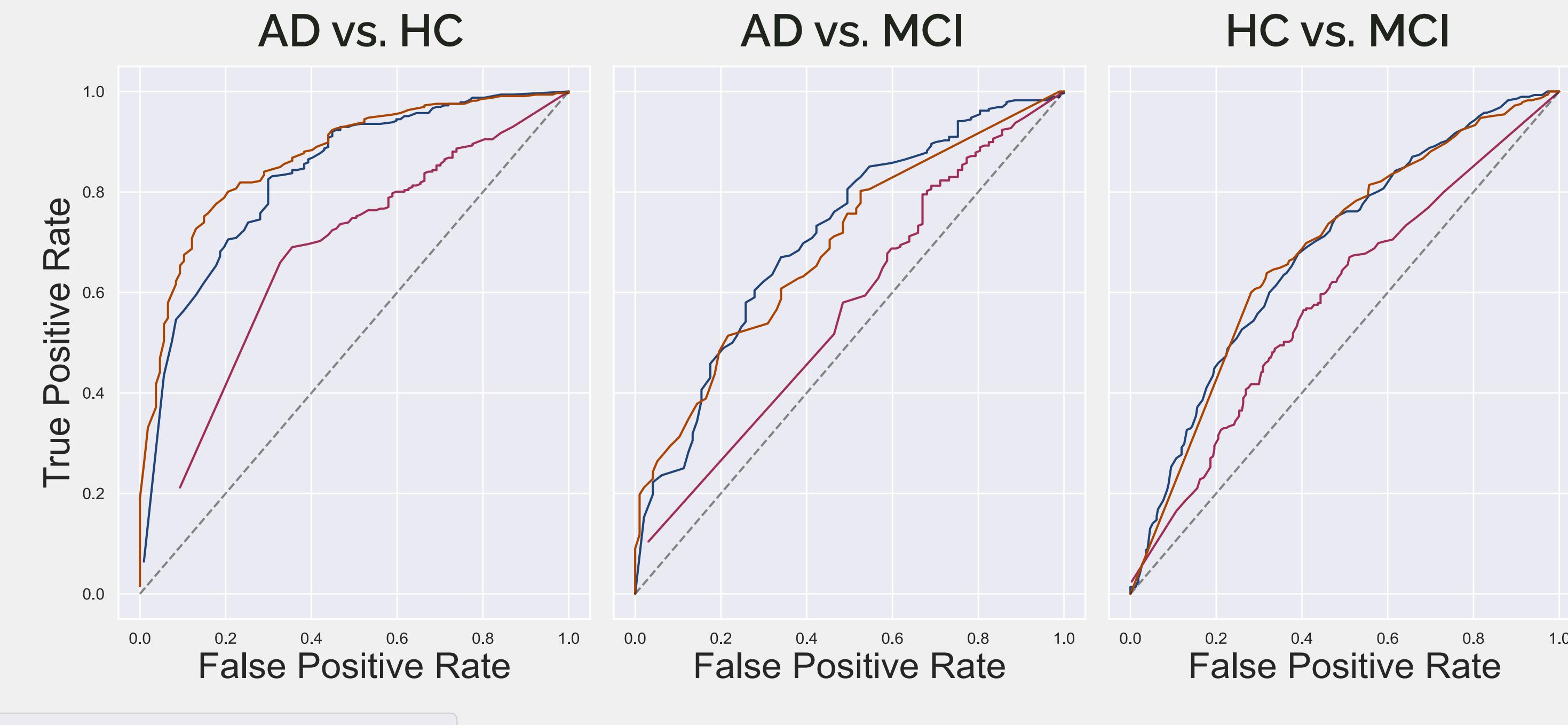


Disease prediction results

Precision-Recall curves



ROC curves

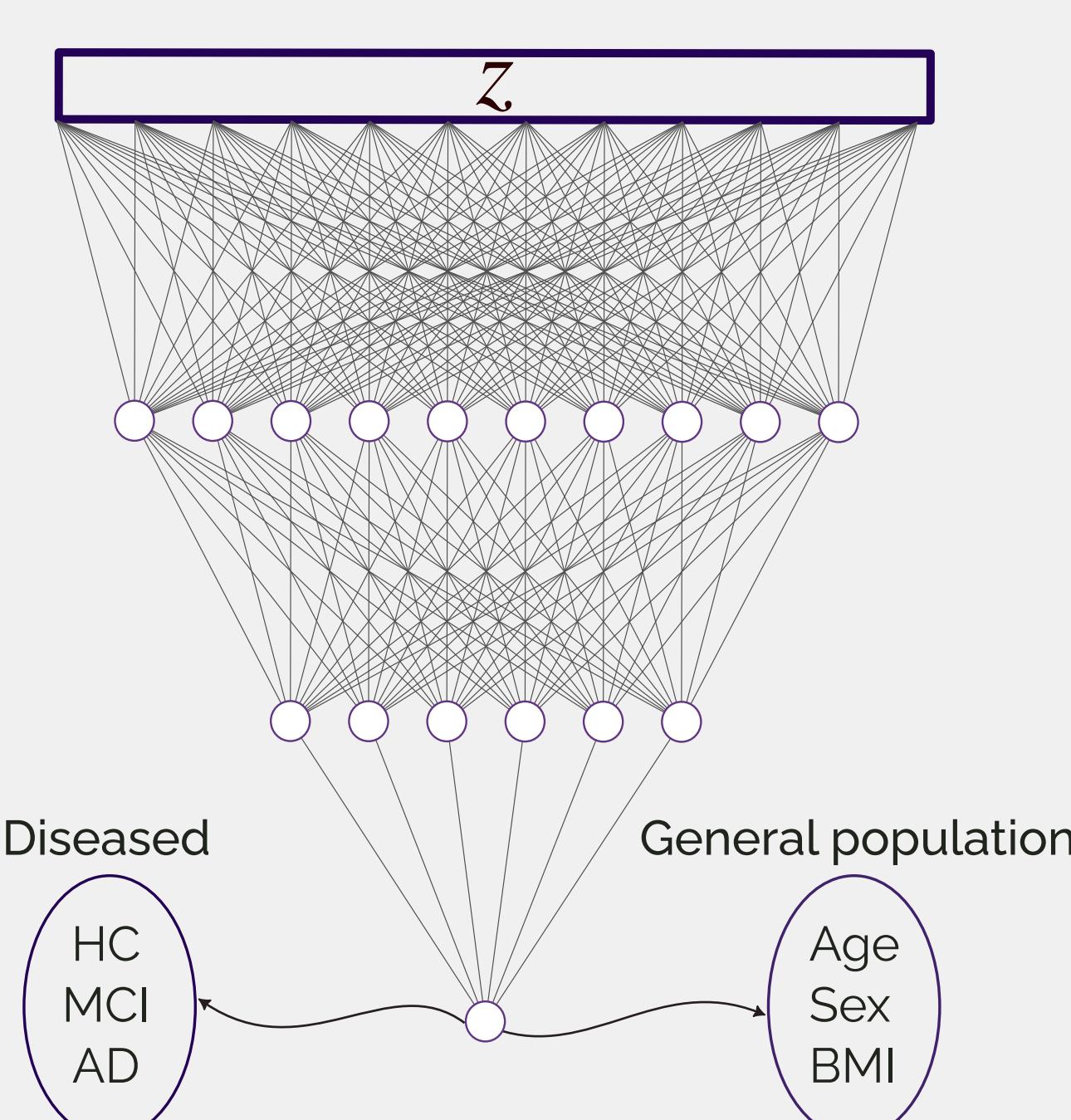


Evaluation

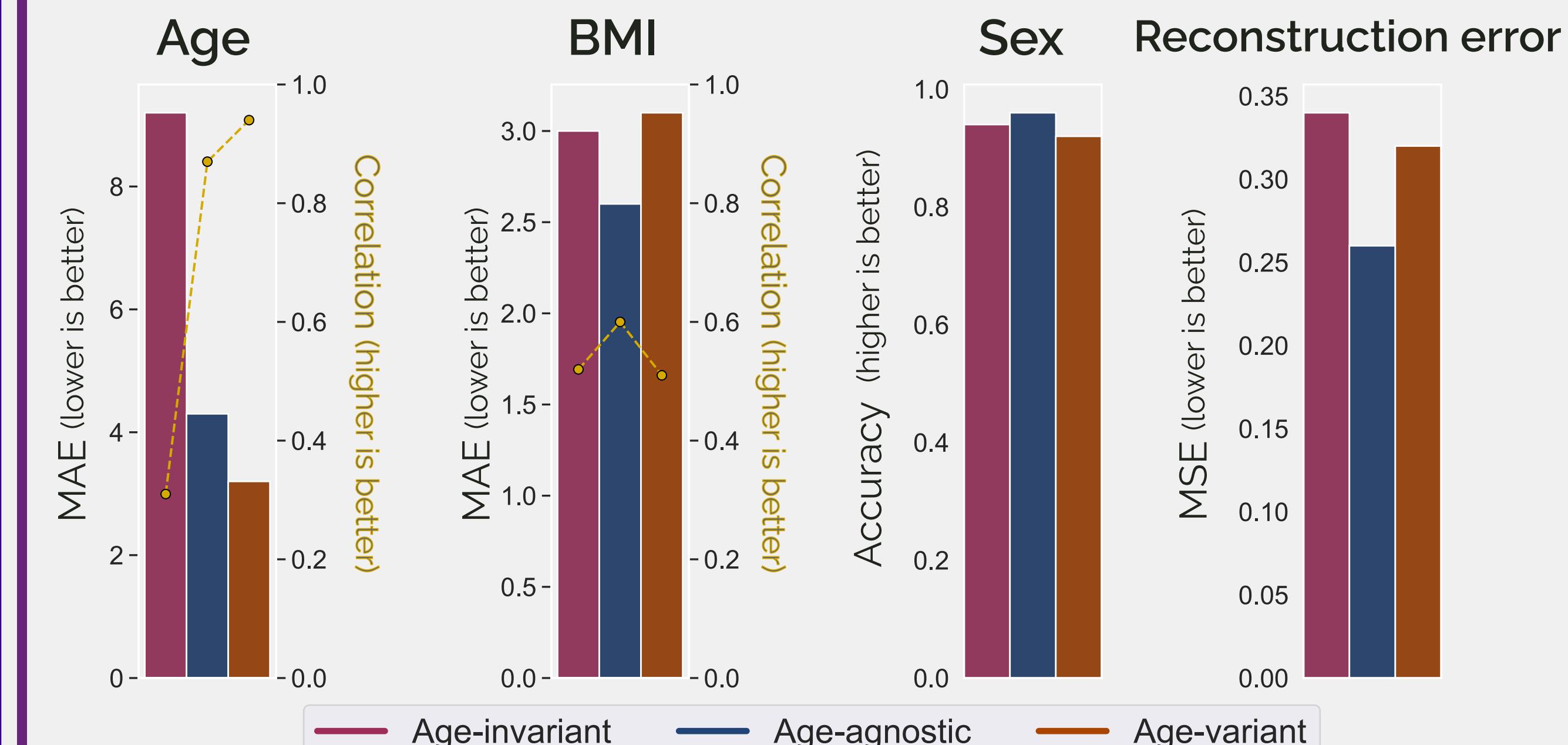
- To test the resulting embeddings, a fully connected neural network with three hidden layers takes the embeddings (z) as input and tries to predict a certain feature.

We train networks on two different tasks:

- Biomarker prediction:** we validate the resulting embeddings by predicting the subjects' age, sex and BMI from the general population validation set.
- Disease prediction:** for each of the three possible pairs of AD, MCI, and HC, we train a network to differentiate between the two.



Biomarker prediction results



Conclusions

- Age-variant embeddings produced the best results for predicting Alzheimer's Disease when compared to Healthy Controls, consistent with the accelerated brain aging hypothesis of neurodegenerative disorders.
- Age-invariant embeddings performed the worst in all cases, probably in part due to its higher reconstruction error, but also suggesting age-related anatomical information is paramount for disease diagnosis.
- Going further, disentangling between age-windows when performing the predictions should help elucidate if accelerated aging is indeed confounding diagnosis.

References

- [1] Kaufmann, T., van der Meer, D., Doan, N.T. et al. (2019). Common brain disorders are associated with heritable patterns of apparent aging of the brain. *Nat Neurosci* 22, 1617–1623.
- [2] Aiskovich, M., Castro, E., Reinen, J., Fadnavis, S., Mehta, A., Li, H., Dhurandhar, A., Cecchi, G., & Polosecki, P. (2024). Fusion of biomedical imaging studies for increased sample size and diversity: a case study of brain MRI. *Frontiers in Radiology*, 4.
- [3] Diederik P Kingma, & Max Welling. (2013). Auto-Encoding Variational Bayes. *arXiv:1312.6114*
- [4] Moyer, D., Gao, S., Brekelmans, R., Steeg, G., & Galstyan, A. (2018). Invariant representations without adversarial training. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems* (pp. 9102–9111).
- [5] Han Peng, Weikang Gong, Christian F. Beckmann, Andrea Vedaldi, & Stephen M. Smith (2021). Accurate brain age prediction with lightweight deep neural networks. *Medical Image Analysis*, 68, 101871.