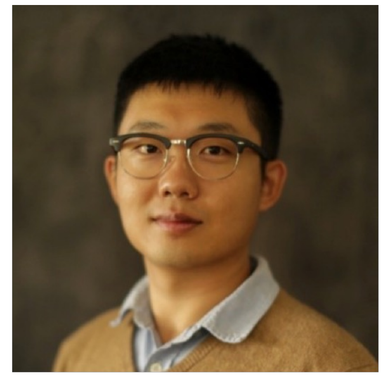
Confounder-Free Continual Learning via Recursive Feature Normalization

Yash Shah [1]

Camila Gonzalez [1]

Mohammad H. Abbasi [1]



Qingyu Zhao [2]

Kilian M. Pohl [1]

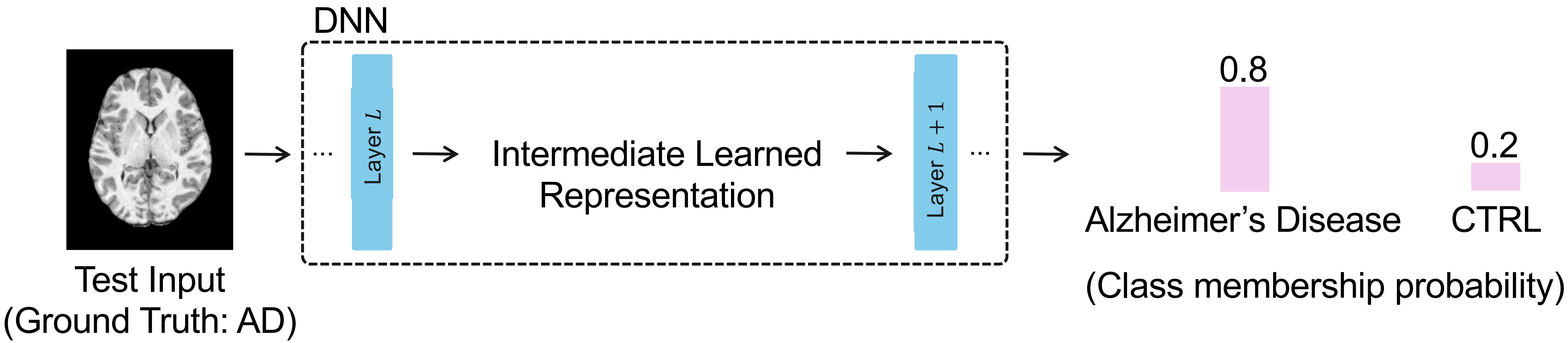
Ehsan Adeli [1]

[1]Stanford University
[2]Weill Cornell Medicine

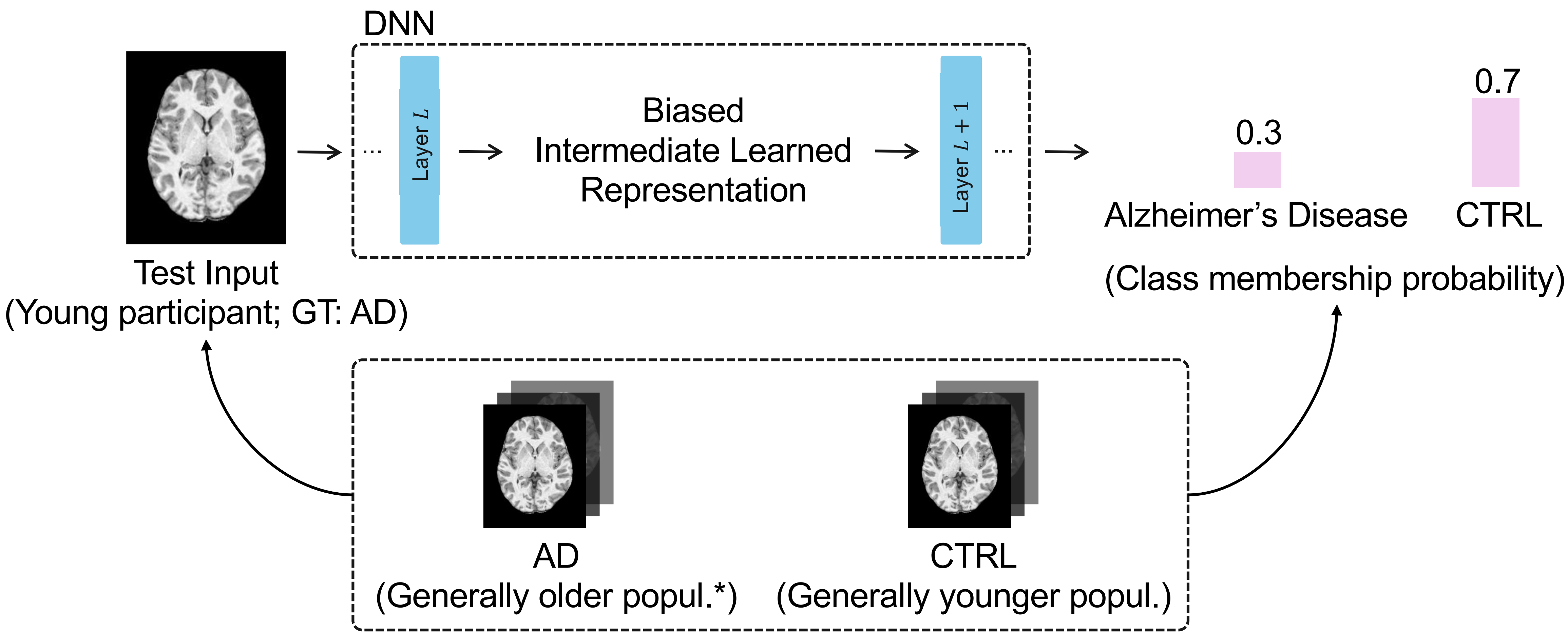
Stanford Translational AI (STAI) Lab
in medicine and mental health

Weill Cornell Medicine

Confounders



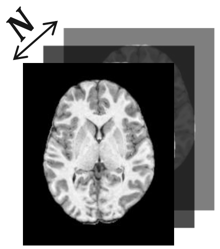
Confounders



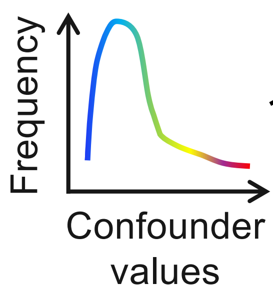
*Rajan et al. *Population estimate of people with clinical Alzheimer's disease and mild cognitive impairment in the United States (2020-2060)*. In Alzheimers Dement, 2021.

Static Learning

Inputs



Distribution of the confounder



Prior Work

[1]*Bias-Resilient Network (BR-Net)*

[2]*Metadata Normalization (MDN)*

[3]*Penalty approach to MDN (PMDN)*

and others...

Confounder-free representation learning

[1]Adeli et al. *Bias-Resilient Neural Network*. Preprint, 2020.
[2]Lu et al. *Metadata Normalization*. In CVPR, 2021.
[3]Vento et al. *A penalty approach for normalizing feature distributions to build confounder-free models*. In MICCAI 2022, vol. 13433 of Lecture Notes in Computer Science.

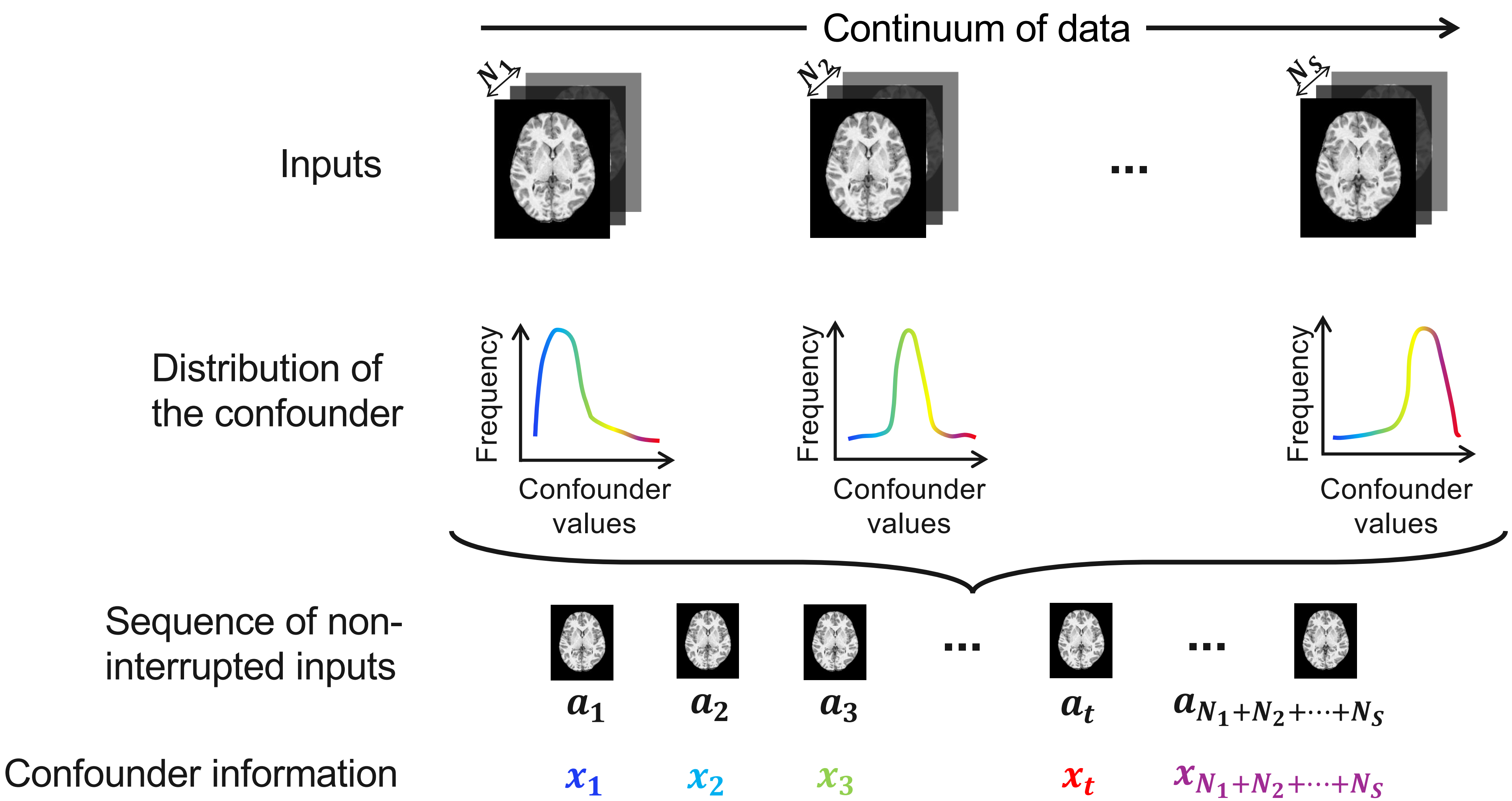
Parameter Updates for MDN*

Say we have N training examples. X is the confounder matrix, and z is the vector of intermediate learned feature representation of the model.

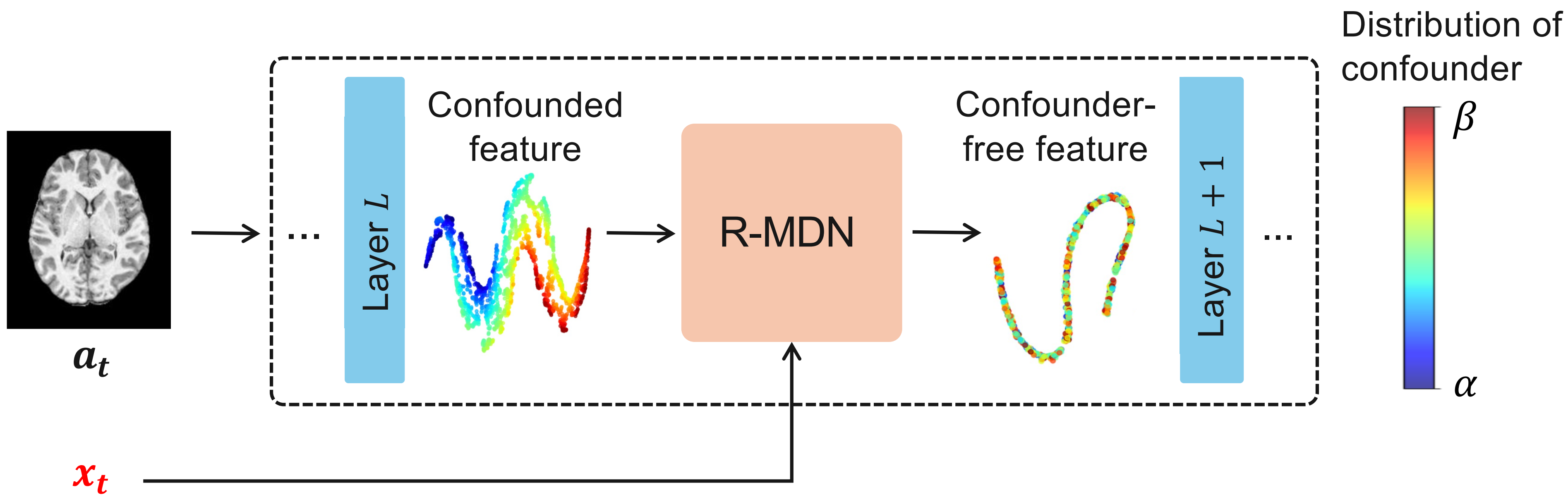
Ordinary Least Squares: $$\beta = \left(\sum_{i=1}^{N} X_i X_i^T\right)^{-1} \left(\sum_{i=1}^{N} z_i X_i\right)$$

*Lu et al. *Metadata Normalization*. In CVPR, 2021.

Continual Learning



Recursive Metadata Normalization

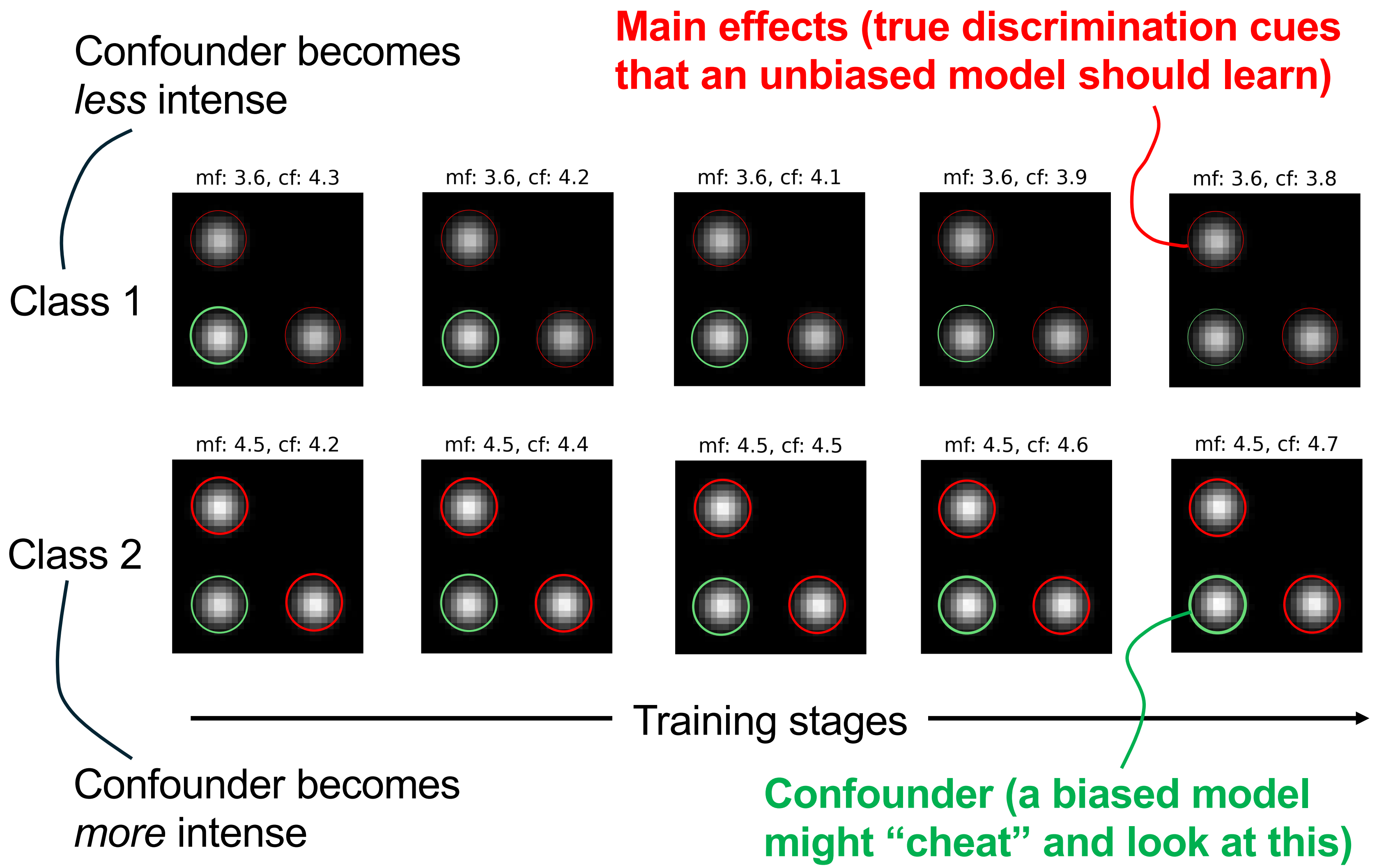


Parameter Updates: $\beta(N+1) = \beta(N) + K(N+1)e(N+1)$,

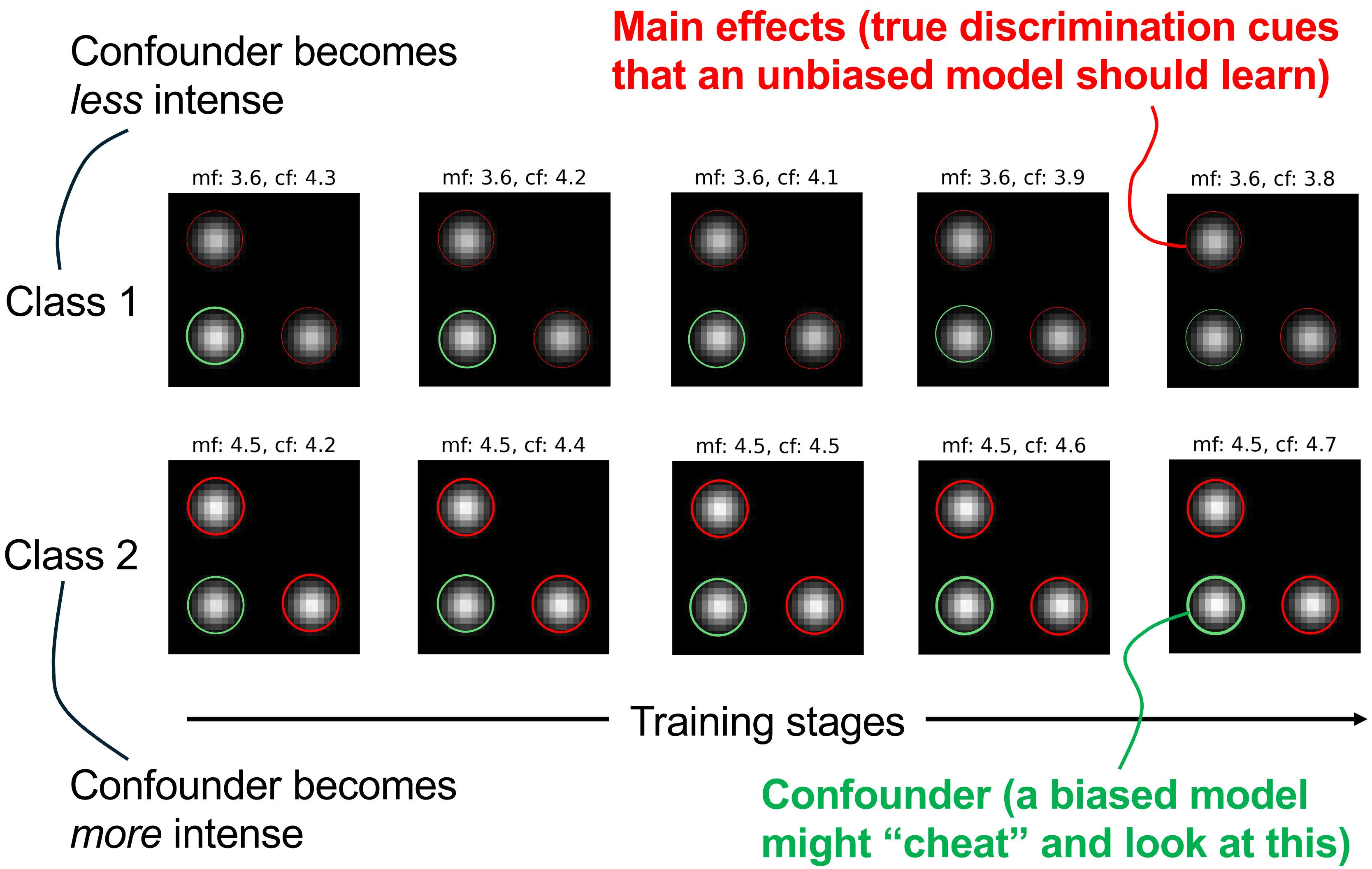
where $e(N+1) = z_{N+1} - X_{N+1}^{T}\beta(N)$ is the a priori error

and $K(N+1)$ is the Kalman Gain at the $N+1$ step

R-MDN effectively removes confounder influence from learned DNN features



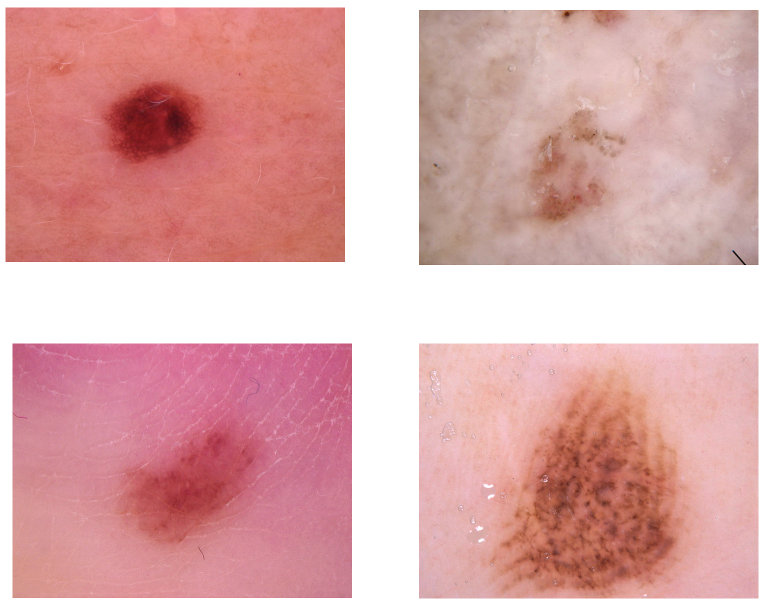
R-MDN effectively removes confounder influence from learned DNN features



Method	Deviation of accuracy from theoretical accuracy (↓)
CNN Baseline	0.18 ± 0.00
BR-Net	0.04 ± 0.03
Stage-specific MDN	0.25 ± 0.00
P-MDN	0.04 ± 0.01
R-MDN	$\mathbf{0.02 \pm 0.01}$

R-MDN is a normalization layer and can be tacked on to various model architectures

Skin lesion classification on HAM10K[1] dataset, with age as confounder



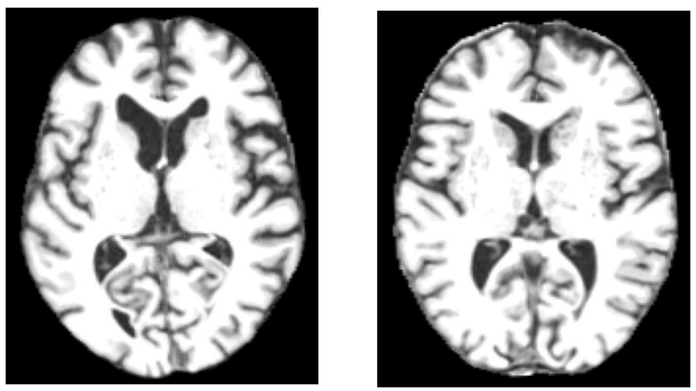
Example images from HAM10K

Method	Accuracy	Average dcor^2 ($\downarrow$)
ViT Baseline	0.7095 ± 0.0626	0.0864 ± 0.0336
BR-Net (Adeli et al., 2020a)	0.7247 ± 0.0627	0.0544 ± 0.0534
P-MDN (Vento et al., 2022)	0.6750 ± 0.0945	0.2595 ± 0.0620
R-MDN (A)	0.5503 ± 0.0541	0.0928 ± 0.0630
R-MDN (B)	0.5288 ± 0.0571	0.0739 ± 0.0555
R-MDN (C)	0.6919 ± 0.0723	**0.0475 ± 0.0247**
EWC (Kirkpatrick et al., 2017)	0.6437 ± 0.0586	0.0938 ± 0.0506
EWC + R-MDN (C)	0.6739 ± 0.0686	0.0592 ± 0.0488
LwF (Li & Hoiem, 2017)	0.7356 ± 0.0757	0.0512 ± 0.0407
LwF + R-MDN (C)	0.7186 ± 0.0736	0.0354 ± 0.0210
PackNet (Mallya & Lazebnik, 2018)	0.6849 ± 0.0745	0.0470 ± 0.0304

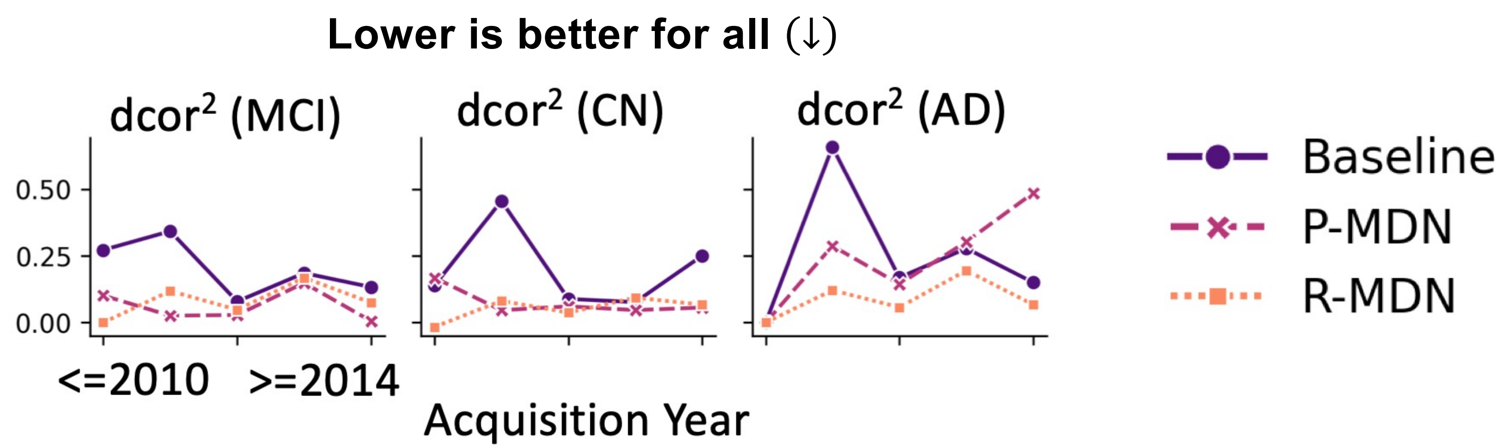
[1]Tschandl et al. *The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions*. In Scientific Data, 2018.

R-MDN can remove the influence from multiple confounders

Diagnostic classification on ADNI[1,2] dataset, with both age and sex as confounders



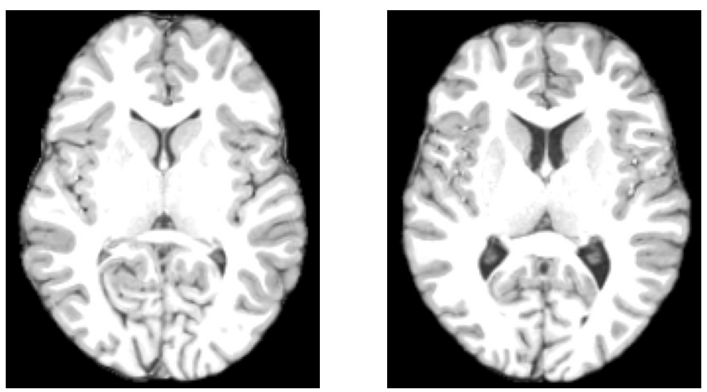
Example images from ADNI



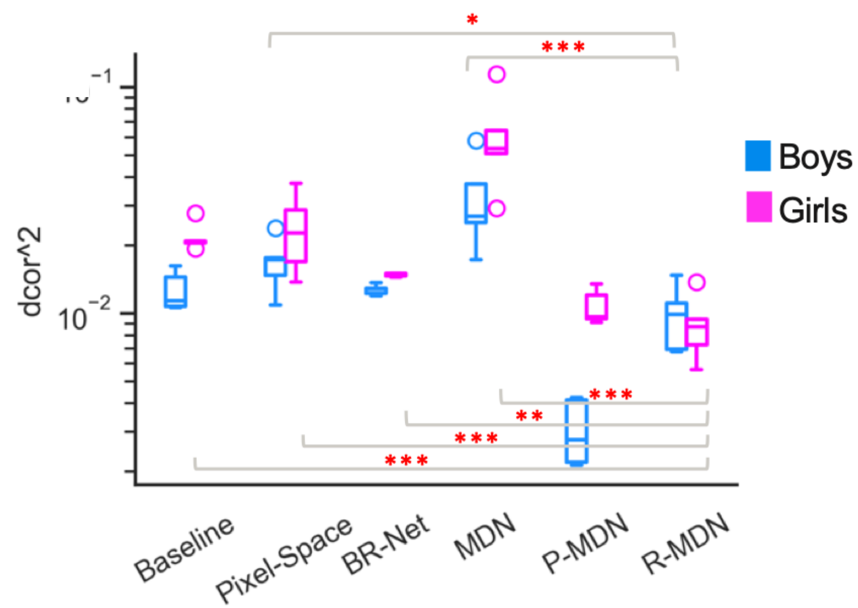
[1]Mueller et al. *The Alzheimer's Disease Neuroimaging Initiative*. In Neuroimaging Clinics of North America, 2005.
[2]Peterson et al. *Alzheimer's Disease Neuroimaging Initiative (ADNI) Clinical Characterization*. In Neurology, 2010.

R-MDN makes equitable predictions across population groups

Sex classification on ABCD[1] dataset, with Pubertal Development Score (PDS) as confounder



Example images from ABCD



[1]Casey et al. *The Adolescent Brain Cognitive Development (ABCD) Study: Imaging Acquisition across 21 sites*. In Developmental Cognitive Neuroscience, 2018.

Interested in knowing more?

ynshah@stanford.edu
eadeli@stanford.edu