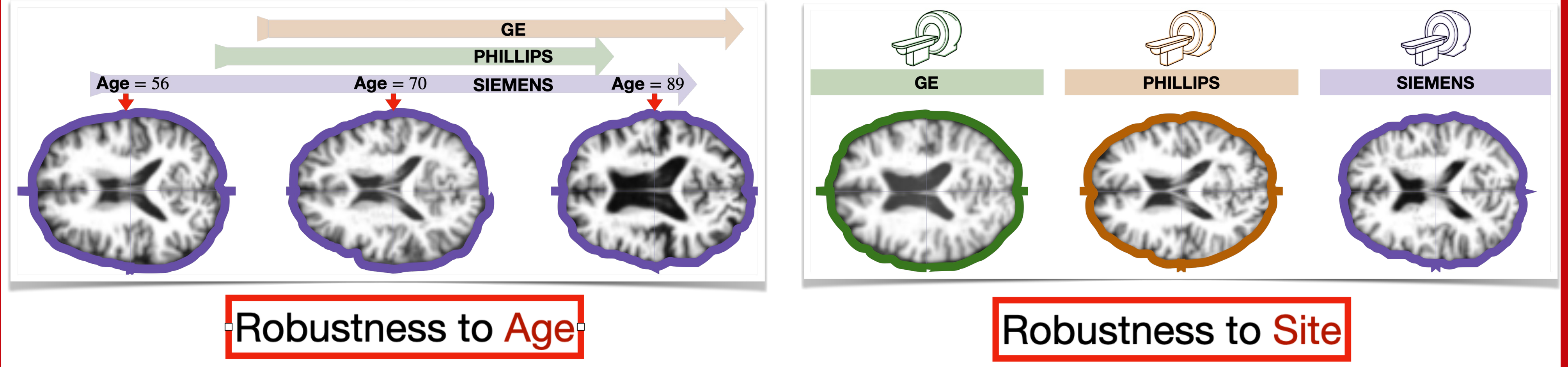
Equivariance Allows Handling Multiple Nuisance Variables When Analyzing Pooled Neuroimaging Datasets

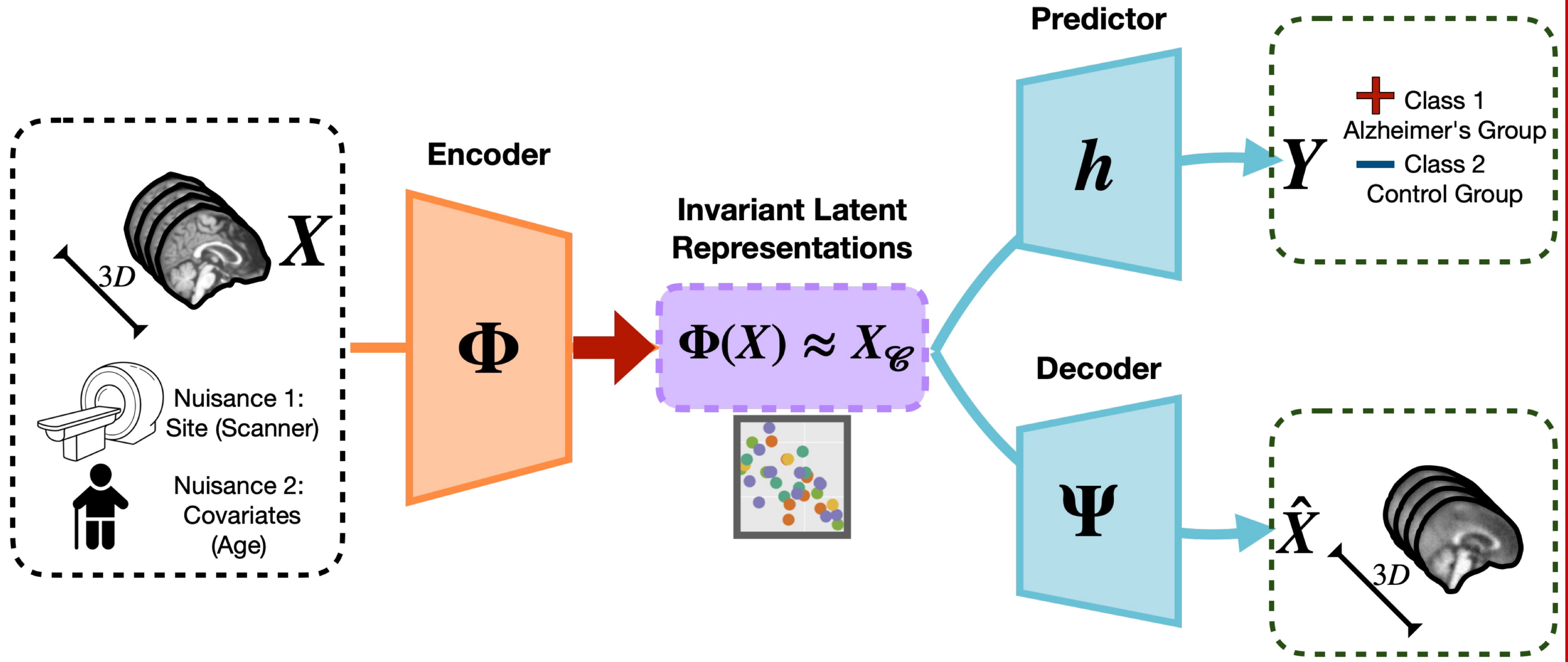
Vishnu Lokhande, Rudrasis Chakraborty, Sathya N. Ravi, Vikas Singh

Multiple Sources of Bias



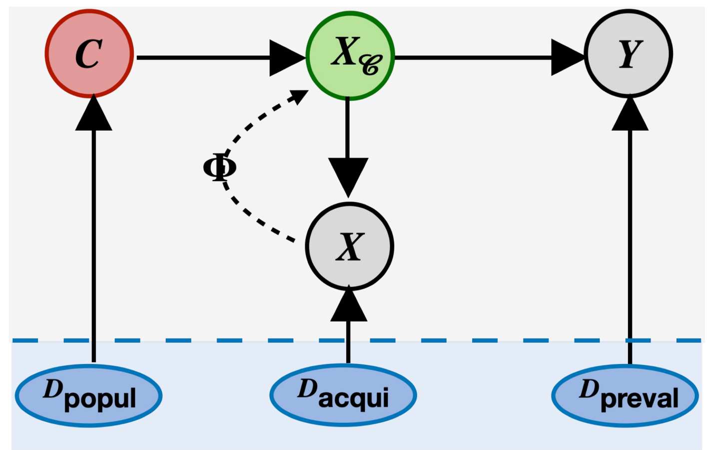
Research Question: Can we pool datasets with **more than one nuisance variables** that concurrently influence data measurements?

Generic Invariant Representation Learning Framework



An **encoder** maps inputs to latent representations. These latent features correspond to the high-level causal representations. Unlike the inputs, **the latent features are robust** to site (scanner) and covariates (age) attributes

Causal Relationship gives Optimization objective

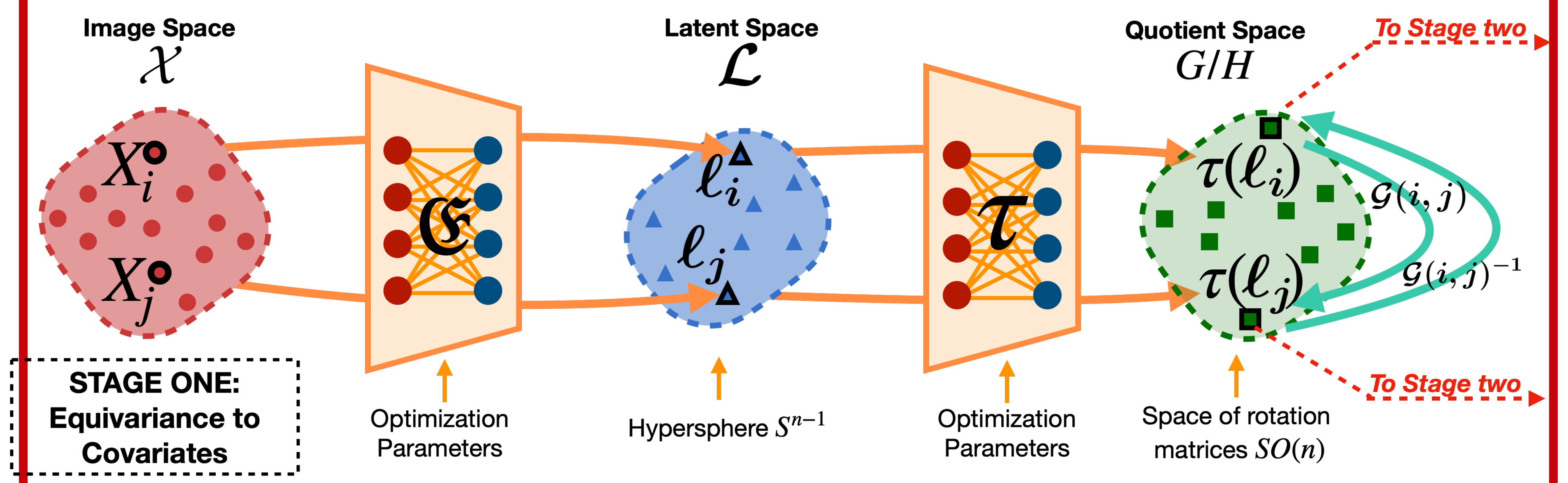


$$X_{\mathscr{C}} \perp D \mid C$$

$$\min_{\Phi} \text{MMD}\left(P_{\text{site}_1}(\Phi(X) \mid C), P_{\text{site}_2}(\Phi(X) \mid C)\right)$$

Invariance Condition: The causal features do not change across sites for samples with the same Covariate value

Stage one - Equivariance to Age



Definition 1: A mapping $f : \mathcal{X} \to \mathcal{Y}$ is said to be G-equivariant iff $f(g \cdot x) = g \cdot f(x) \quad g \in \text{Group } G$

Stage one - We learn a mapping to a space providing flexibility to characterize changes in covariates as a group action.

Stage two - Invariance to Site

Step 1: Preserve equivariance from stage one with an equivariant mapping

Lemma 5: For a mapping $\tau : \mathcal{L} \to G/H$ and arbitrary mapping $b : \mathcal{L} \to Z$ $\Phi(\ell) = \tau(\ell) \cdot b(\tau(\ell)^{-1} \cdot \ell)$ is G-equivariant i.e, $\Phi(g \cdot \ell) = g\Phi(\ell)$

Step 2: Optimize MMD loss alongside reconstruction and prediction losses

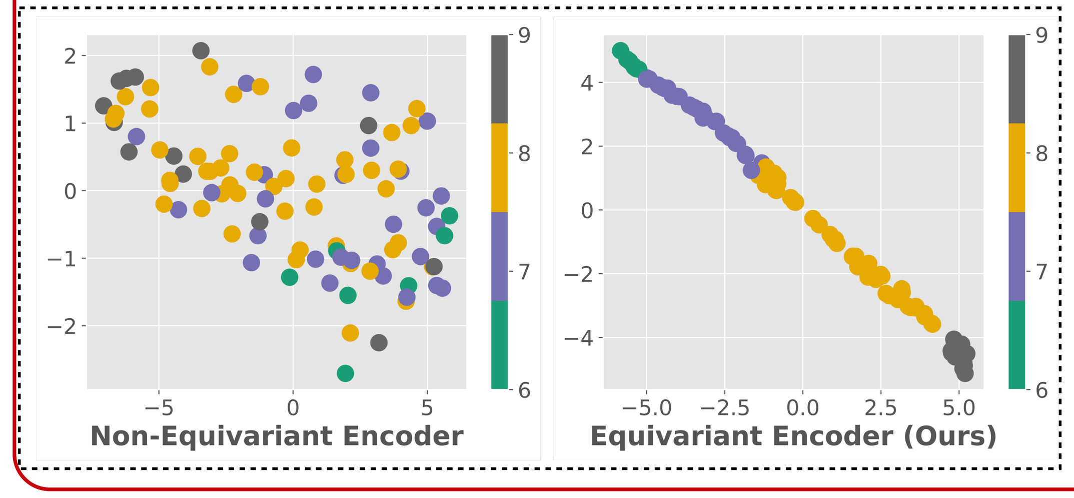
$$L_{\text{stage}_2} = \sum \underbrace{\|\ell - \Psi(\Phi(\ell))\|^2}_{\text{Reconstruction loss}} + \underbrace{\|Y - h(\Phi(\ell))\|^2}_{\text{Prediction loss}} + \text{MMD}$$

subject to $\Phi(\ell) = \tau(\ell) \cdot b\left(\tau(\ell)^{-1} \cdot \ell\right)$

Stage two - We learn a second encoding to a generic vector space by first ensuring that the equivariance properties from Stage one are preserved. Such an encoding is then tuned to achieve invariance to site.

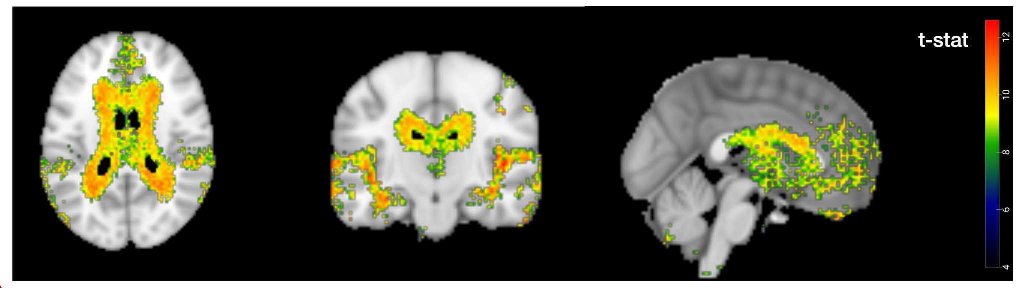
Experiments

Latent representations t-SNE plots



An equivariant encoder ensures that the latent features are evenly distributed and bear a monotonic trend with respect to the changes in the age covariate value

Statistical Analysis on the reconstructed outputs



A high density of significant voxels indicates that our method preserves disease related signal after pooling across scanners

Distribution of age covariate in the ADNI dataset

Despite the mismatch of support across scanner attributes, our approach minimizes the MMD measure (desirable)

References

1. Elias Bareinboim and Judea Pearl. Causal inference and thedata-fusion problem .Proceedings of the National Academy of Sciences, 113(27):7345–7352, 2016
2. Benjamin Bloem-Reddy and Yee Whye Teh. Probabilistics ymmetries and invariant neural networks. Journal of Ma-chine Learning Research, 21(90):1–61, 2020
3. Daniel C Castro, Ian Walker, and Ben Glocker. Causal-ity matters in medical imaging. Nature Communications,11(1):1–10, 2020
4. Divyat Mahajan, Shruti Tople, and Amit Sharma. Domain generalization using causal matching. In International Con-ference on Machine Learning, pages 7313–7324. PMLR,2021

Checkout our GitHub page