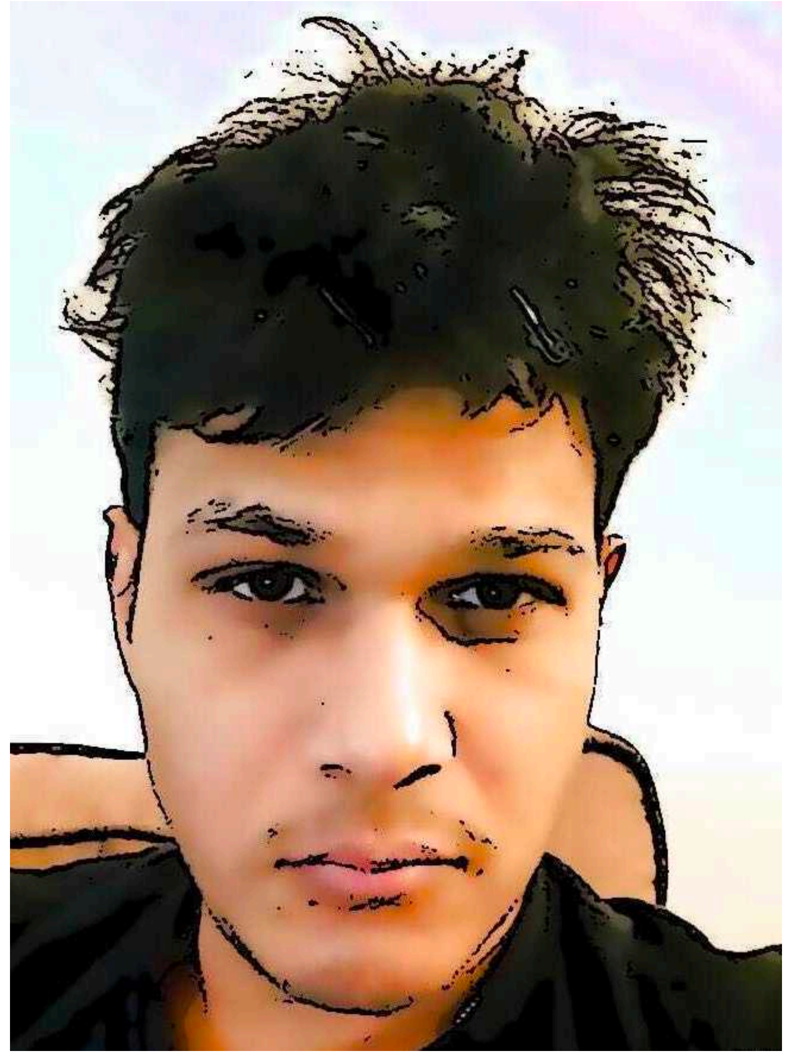
CVPR JUNE 19-24 2022 NEW ORLEANS • LOUISIANA

Equivariance Allows Handling Multiple Nuisance Variables When Analyzing Pooled Neuroimaging Datasets

Presenter: Vishnu Lokhande
Joint Work with Rudrasis Chakraborty, Sathya Ravi and Vikas Singh

Vishnu Lokhande



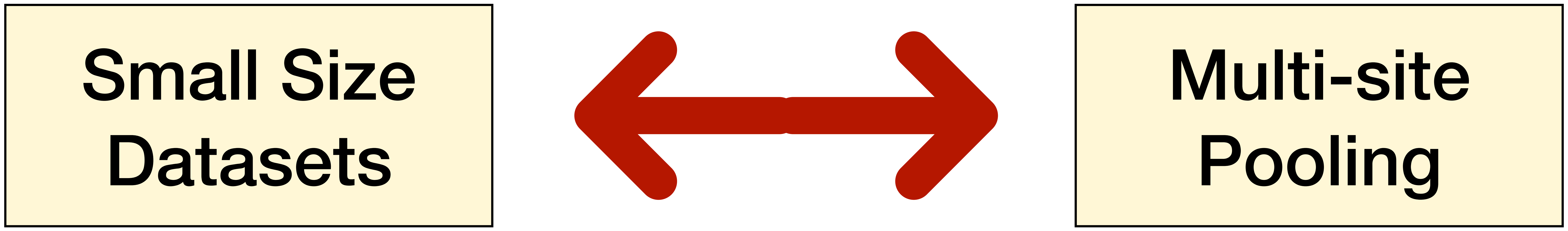
Rudra Chakraborty

Sathya Ravi

Vikas Singh

Dataset Pooling in Neuroimaging studies

Dataset Pooling in Neuroimaging studies



Dataset Pooling in Neuroimaging studies

Small Size Datasets

Multi-site Pooling

Improves statistical power: Better disease-pattern identification

Dataset Pooling in Neuroimaging studies

Small Size Datasets

Multi-site Pooling

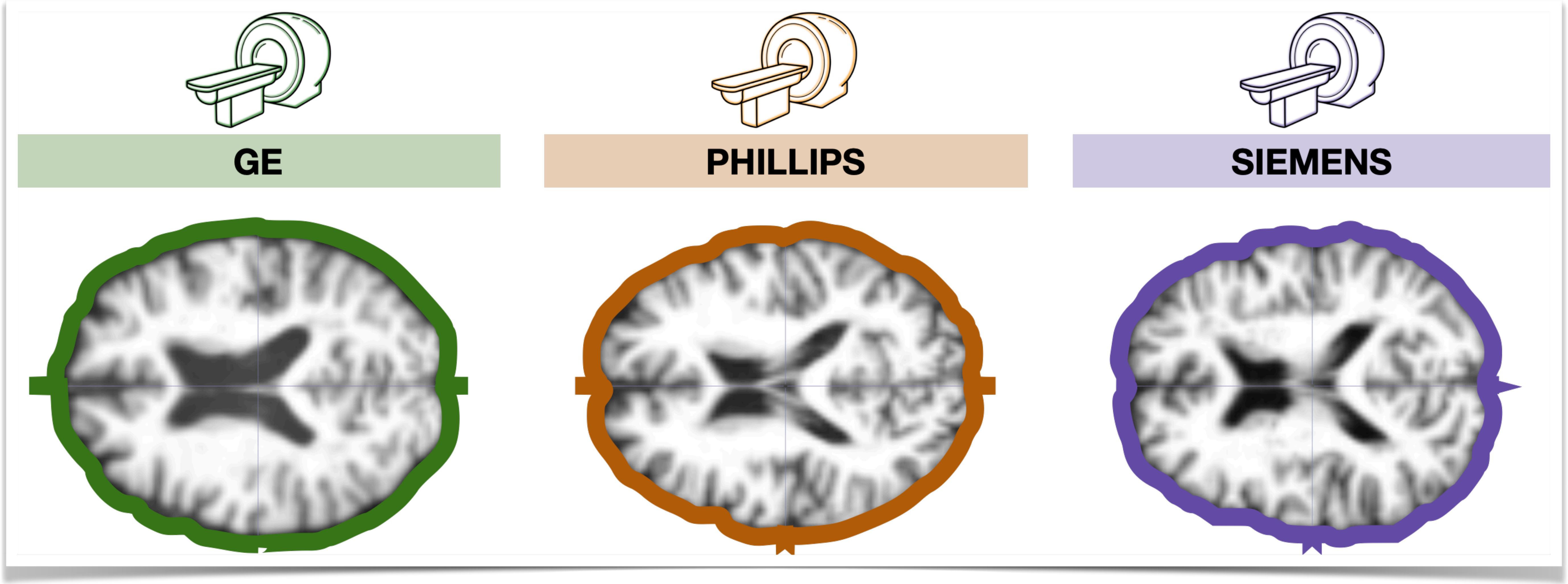
Improves statistical power: Better disease-pattern identification

Models impacted with site-specific nuisance attributes

Nuisance Attribute types

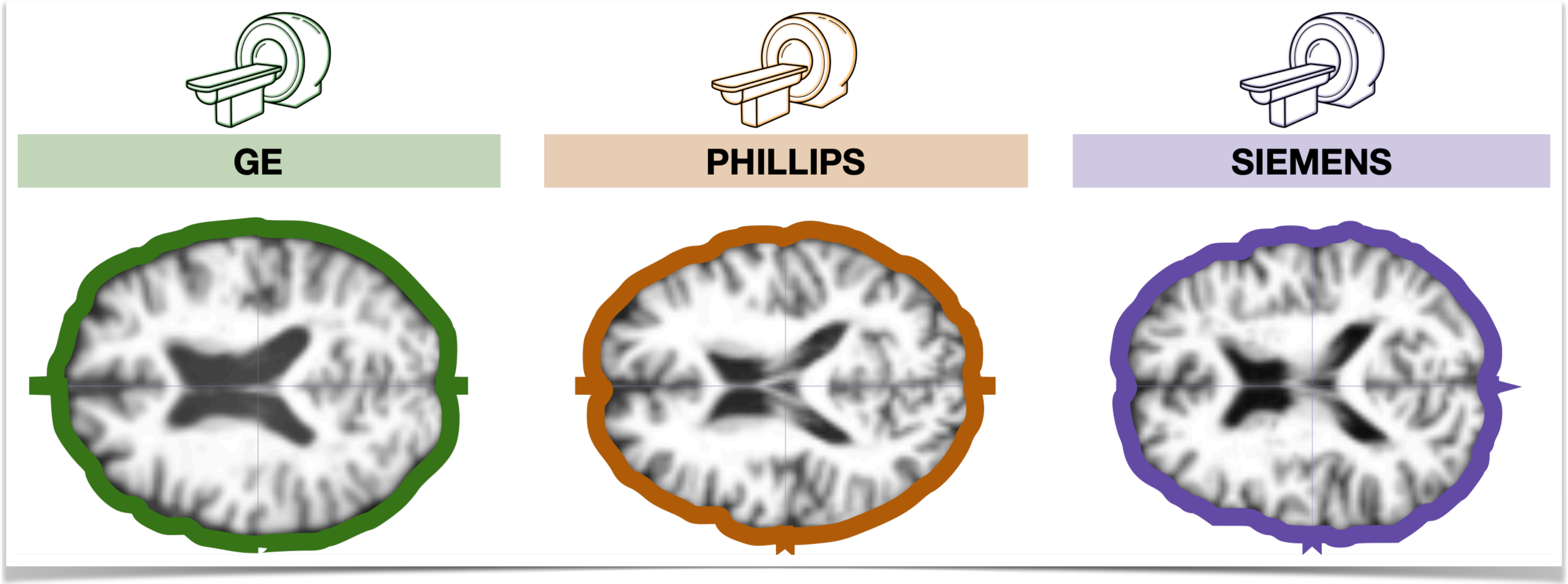
Nuisance Attribute types

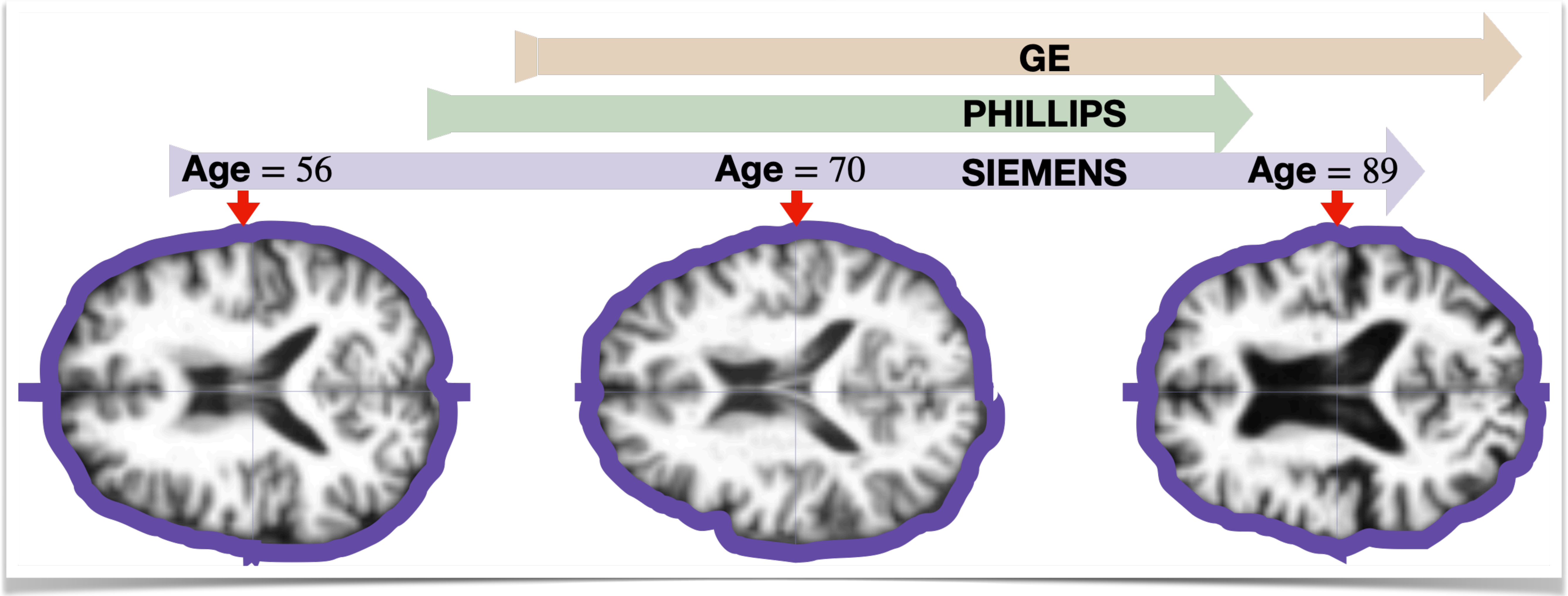
Data Acquisition Bias



Nuisance Attribute types

Data Acquisition Bias





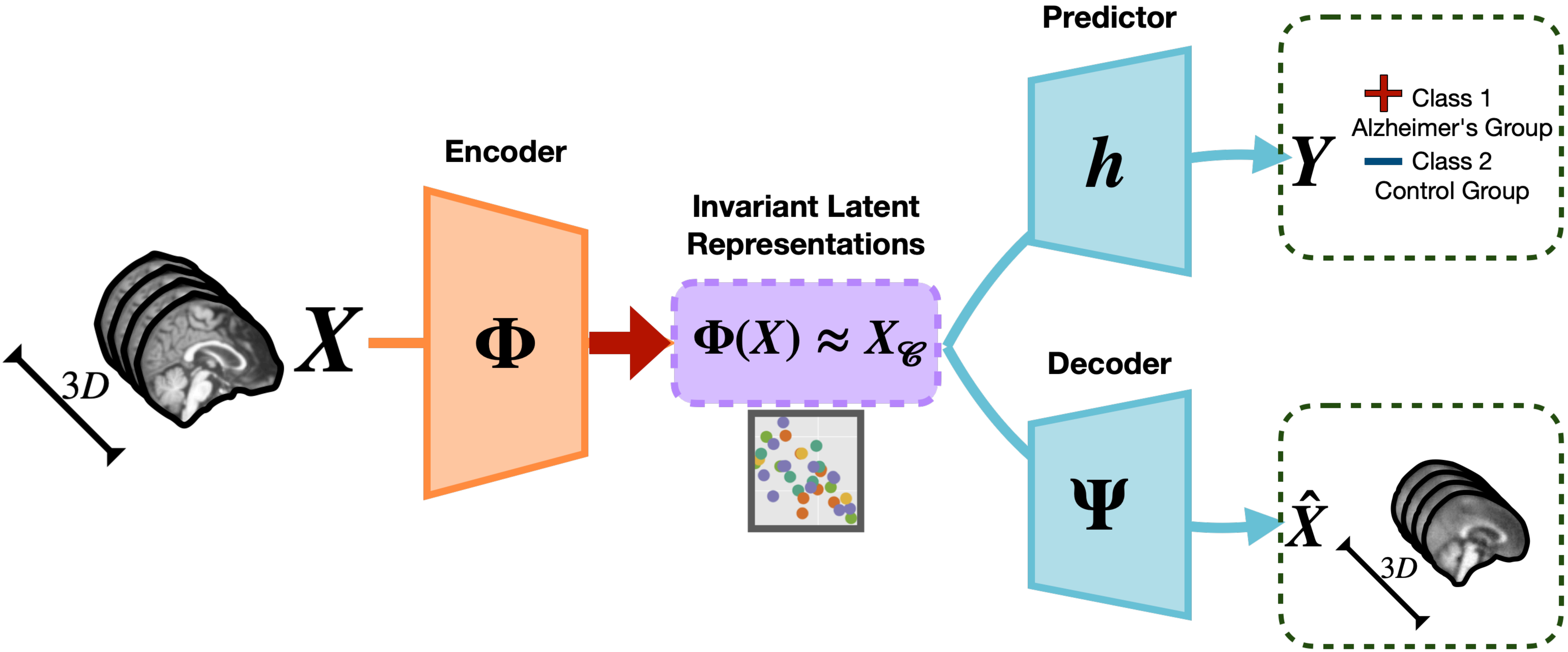
Population Bias

Research Question

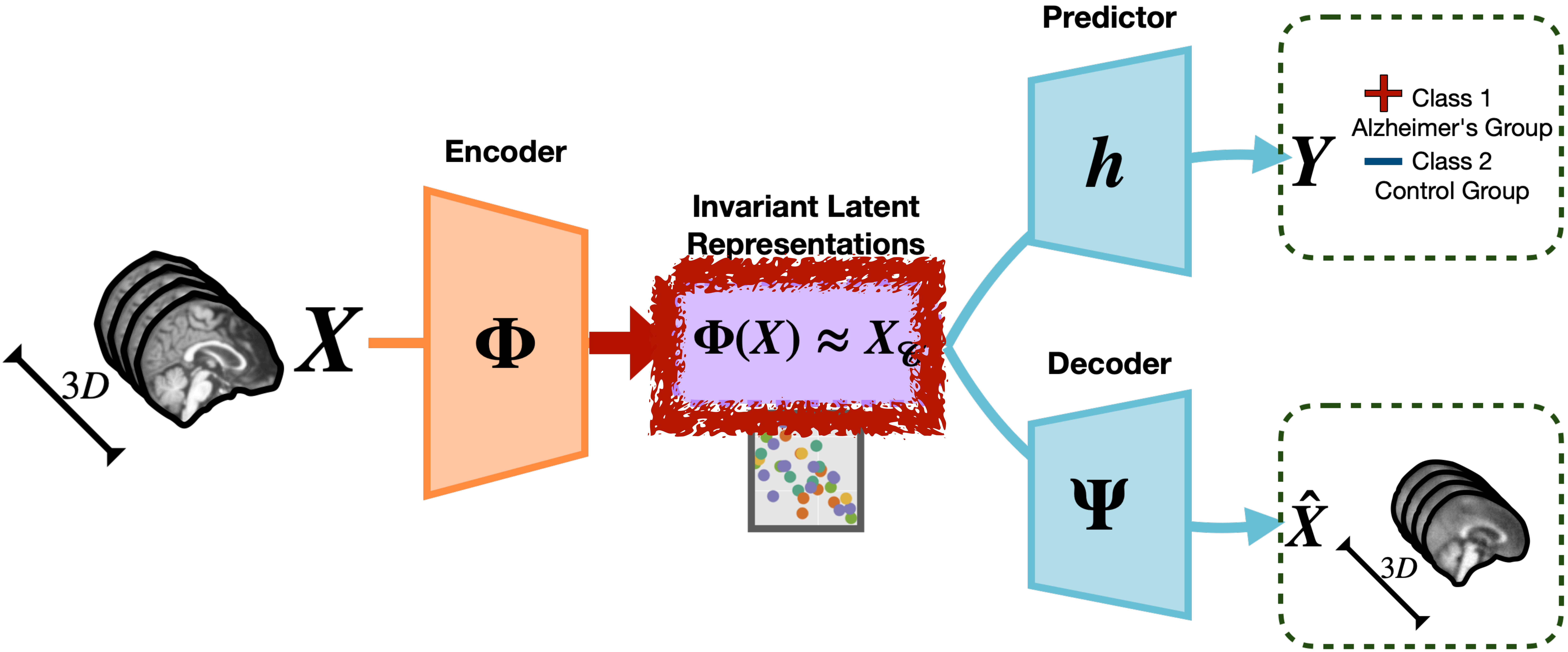
Can we pool datasets in the presence of more than one nuisance attributes ?

Generic Invariant Representation Learning Framework

Generic Invariant Representation Learning Framework



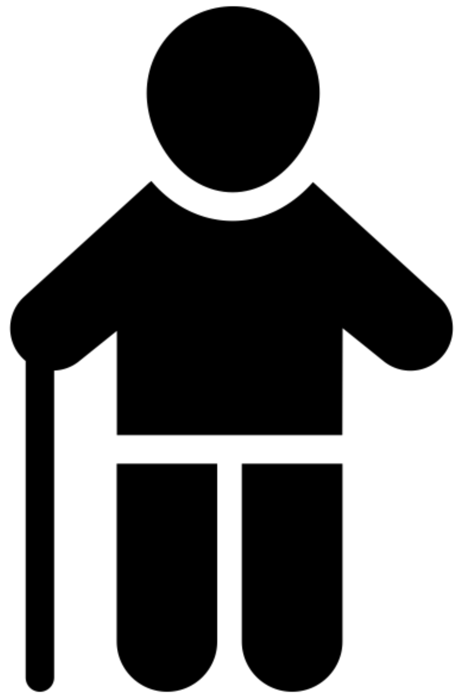
Generic Invariant Representation Learning Framework



Problems with Multiple Nuisance Attributes

Problems with Multiple Nuisance Attributes

Nuisance 1:
Site (Scanner)



Nuisance 2:
Covariates (Age)

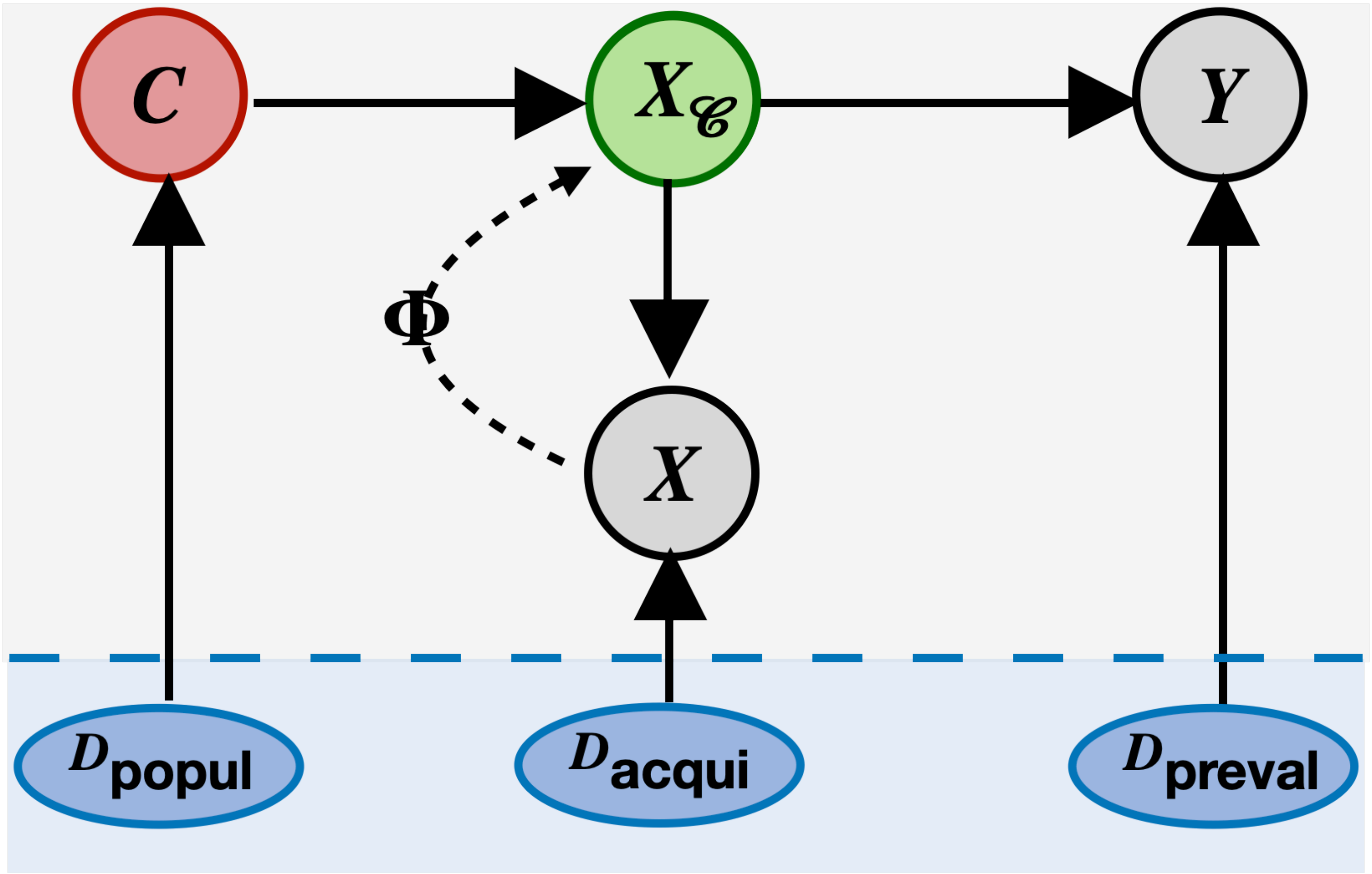
Closest Match of participants across sites

Discards Samples outside the common support

Causal Diagram + Invariance Condition = Optimization objective

Causal Diagram + Invariance Condition = Optimization objective

Causal Diagram



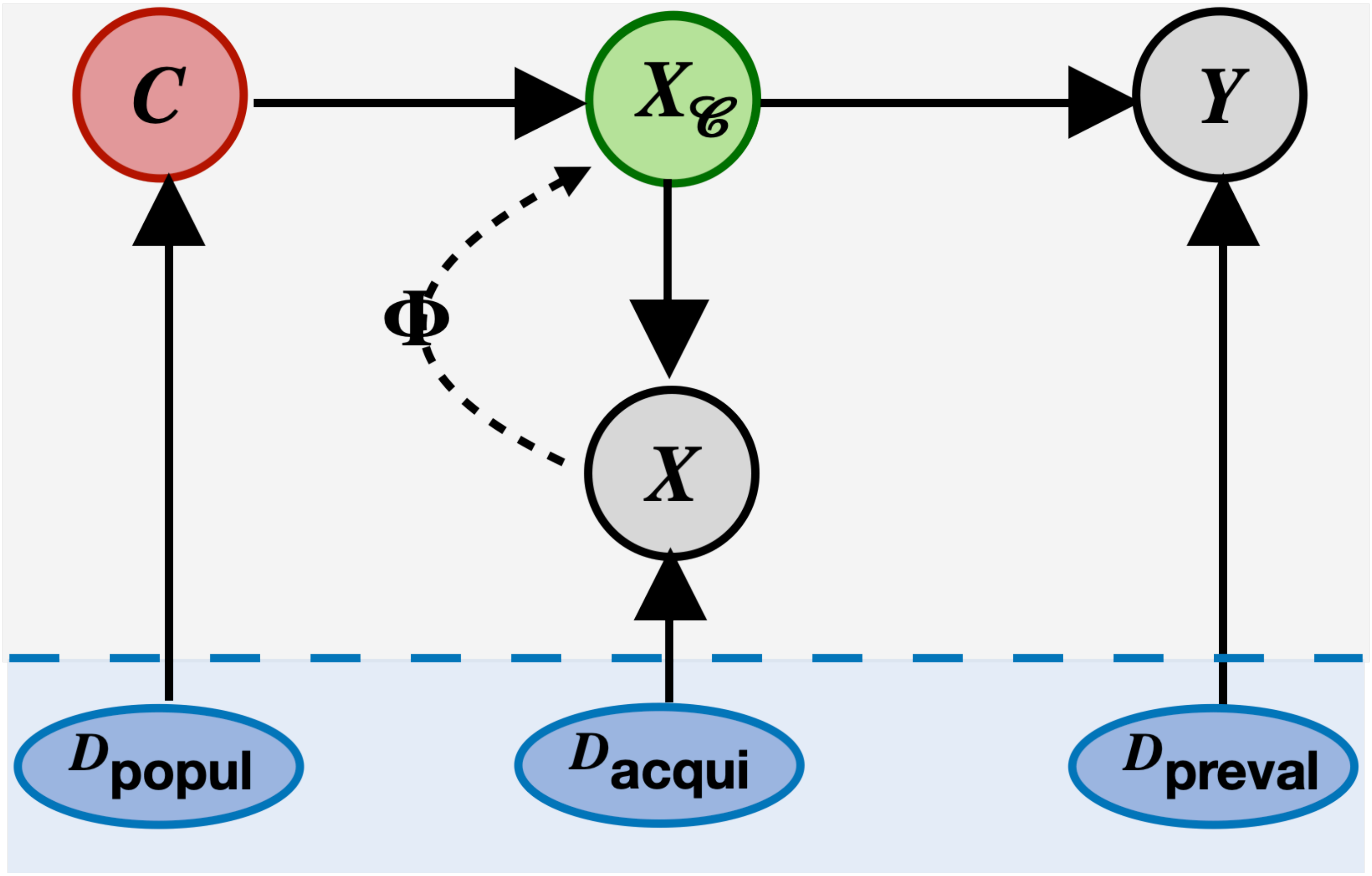
Invariance Condition

$X_{\mathscr{C}}$ does not change across the sites for the **same value** of C

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

Causal Diagram + Invariance Condition = Optimization objective

Causal Diagram



Invariance Condition

$X_{\mathscr{C}}$ does not change across the sites for the **same value** of C

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

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

Improved Distribution Matching via Equivariant Mappings

Improved Distribution Matching via Equivariant Mappings

Equivariant Mappings!

Improved Distribution Matching via Equivariant Mappings

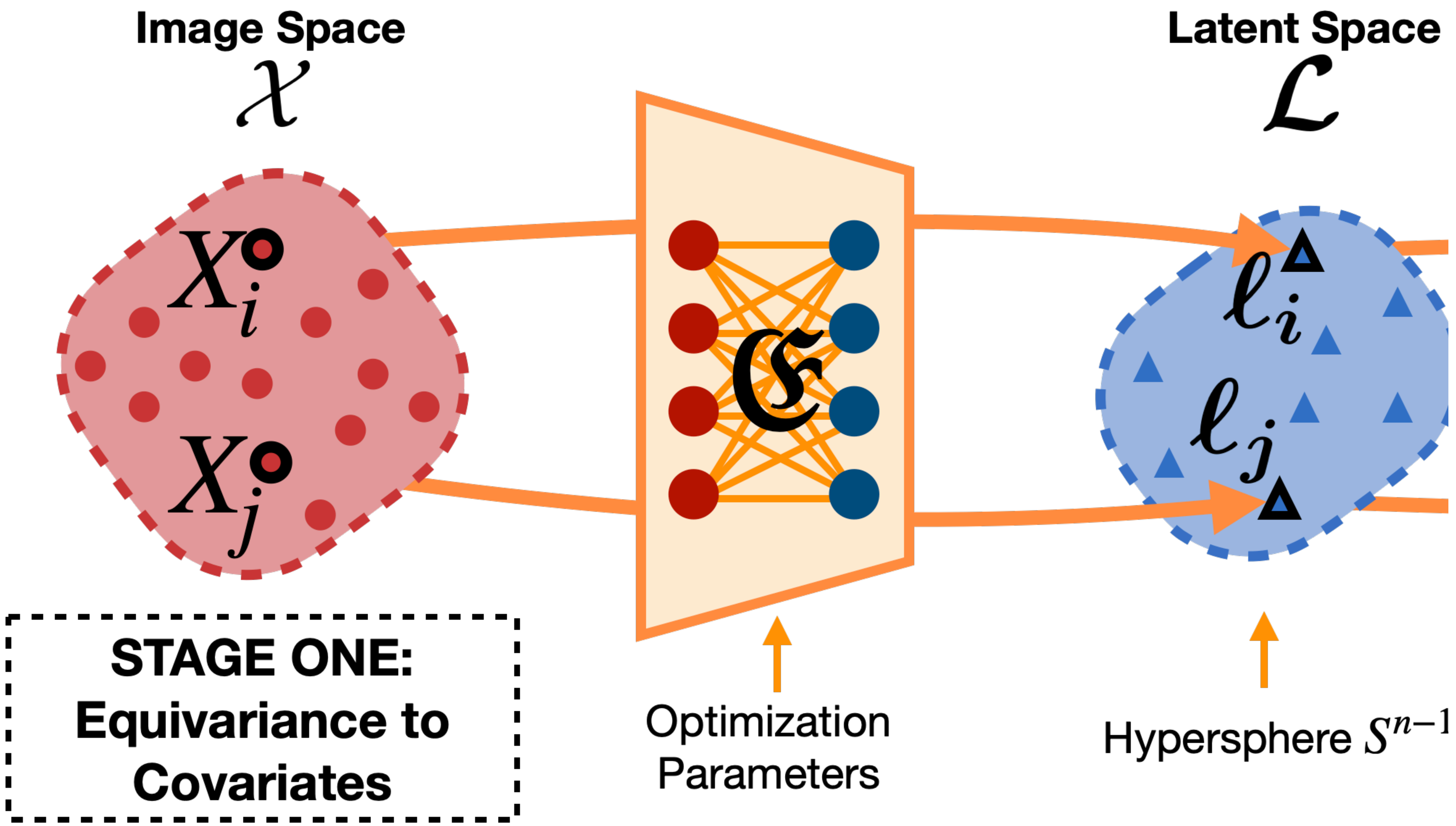
Equivariant Mappings!

Definition 1: A mapping $f: \mathscr{X} \to \mathscr{Y}$ is said to be G-equivariant iff

$$f(g \cdot x) = g \cdot f(x) \quad g \in \text{Group } G$$

Stage one - Equivariance to Age

Stage one - Equivariance to Age



Stage one - Equivariance to Age

Stage two - Invariance to Site

Stage two - Invariance to Site

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

Lemma 5: For a mapping $\tau : \mathscr{L} \to G/H$ and arbitrary mapping $b : \mathscr{L} \to Z$

$\Phi(\ell) = \tau(\ell) \cdot b(\tau(\ell)^{-1} \cdot \ell)$ is G-equivariant i.e, $\Phi(g \cdot \ell) = g\Phi(\ell)$

Stage two - Invariance to Site

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

Lemma 5: For a mapping $\tau : \mathscr{L} \to G/H$ and arbitrary mapping $b : \mathscr{L} \to Z$

$\Phi(\ell) = \tau(\ell) \cdot b(\tau(\ell)^{-1} \cdot \ell)$ is G-equivariant i.e, $\Phi(g \cdot \ell) = g\Phi(\ell)$

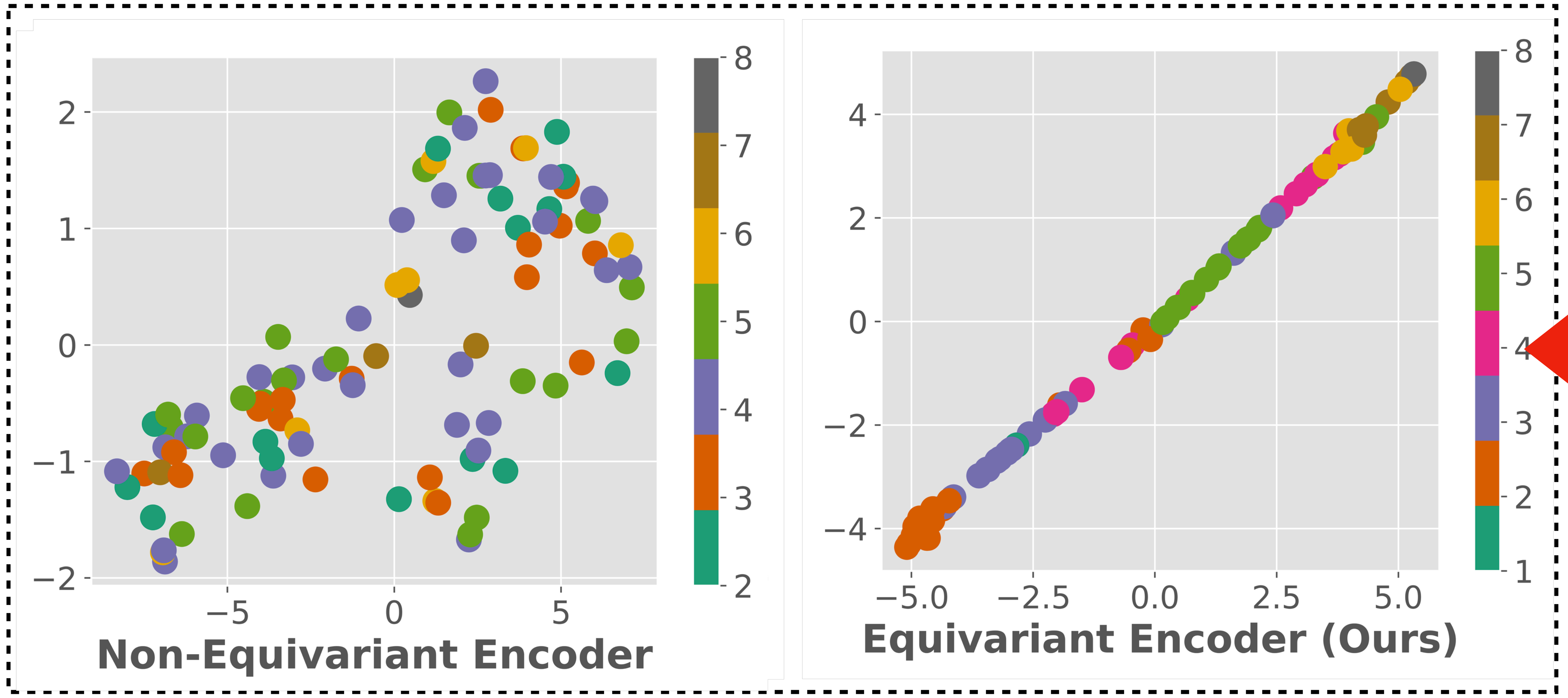
Step 2: Optimize MMD loss alongside reconstruction and prediction losses

$$L_{\text{stage2}} = \sum \underbrace{\boxed{\|\ell - \Psi(\Phi(\ell))\|^2}}_{\text{Reconstruction loss}} + \underbrace{\boxed{\|Y - h(\Phi(\ell))\|^2}}_{\text{Prediction loss}} + \text{MMD}$$

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

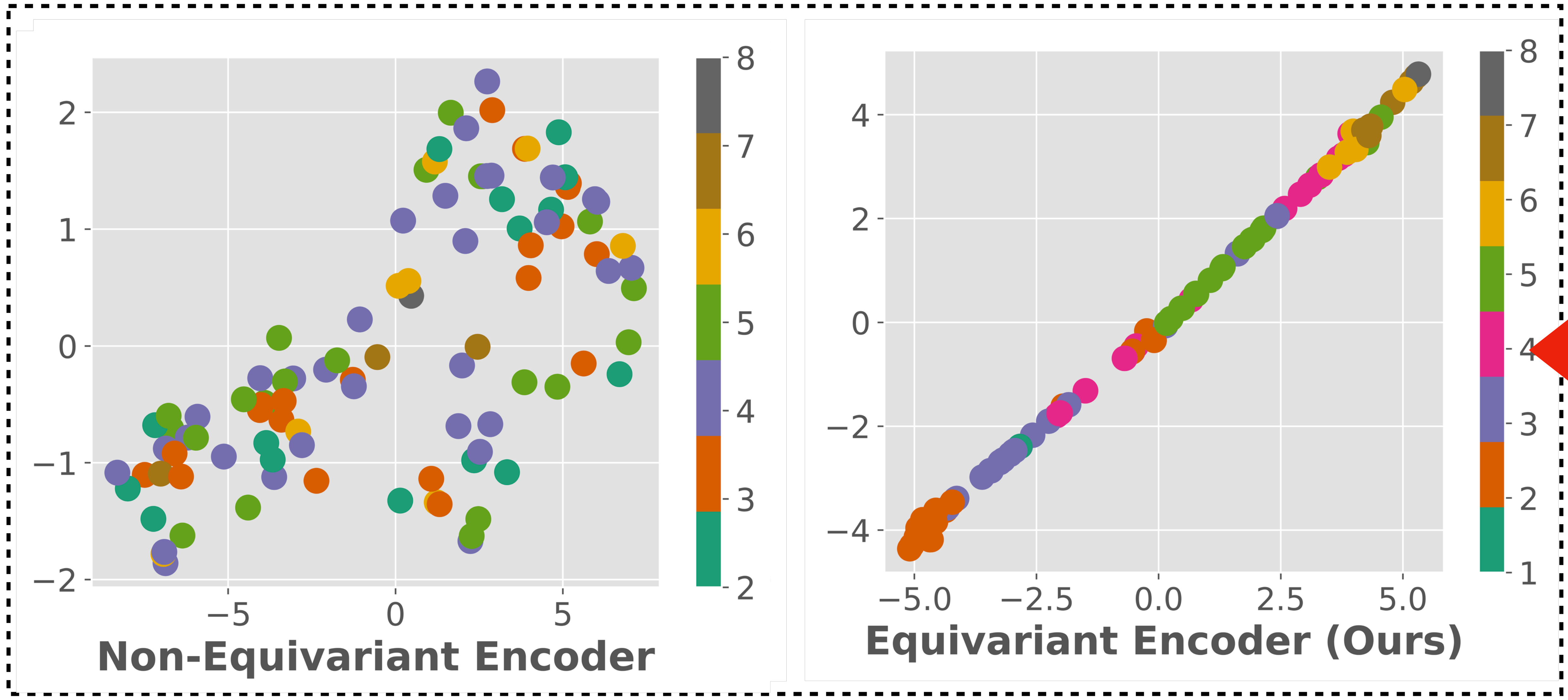
Experimental Results

Experimental Results



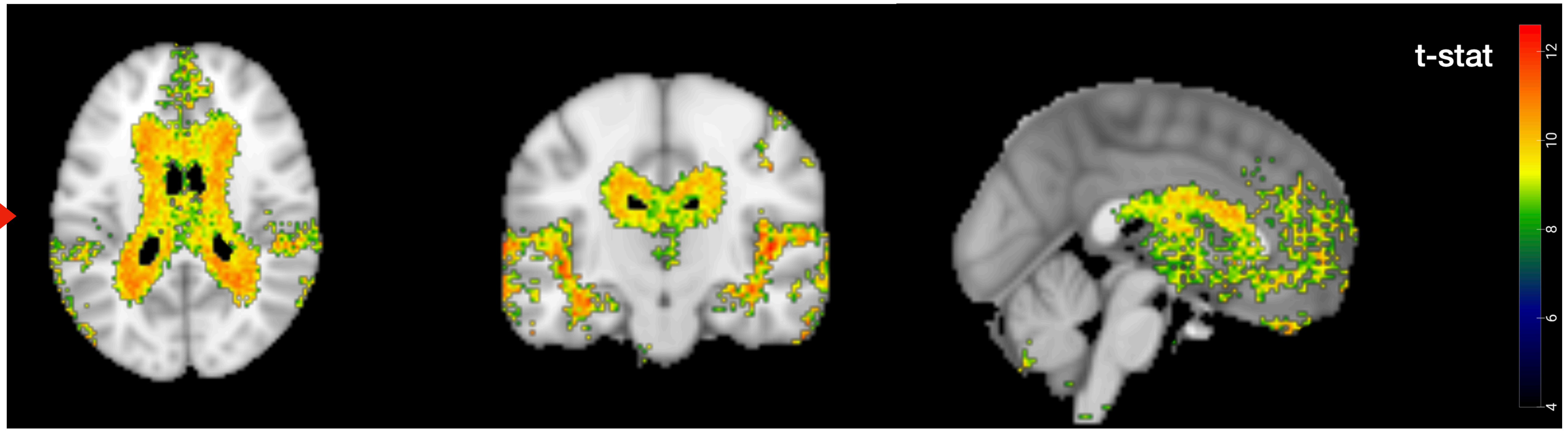
Equivariant encoder results in a monotonic trend in the latent features as covariate values are varied.

Experimental Results



Equivariant encoder results in a monotonic trend in the latent features as covariate values are varied.

Preserves disease-specific signal as indicated by high density of significant voxels



Conclusion

Conclusion

A Method to

- ✓ Pool datasets from multiple sites
- ✓ Data influenced by multiple nuisance attributes
- ✓ Distribution of covariates not identical

Conclusion

A Method to

- Pool datasets from multiple sites
- Data influenced by multiple nuisance attributes
- Distribution of covariates not identical