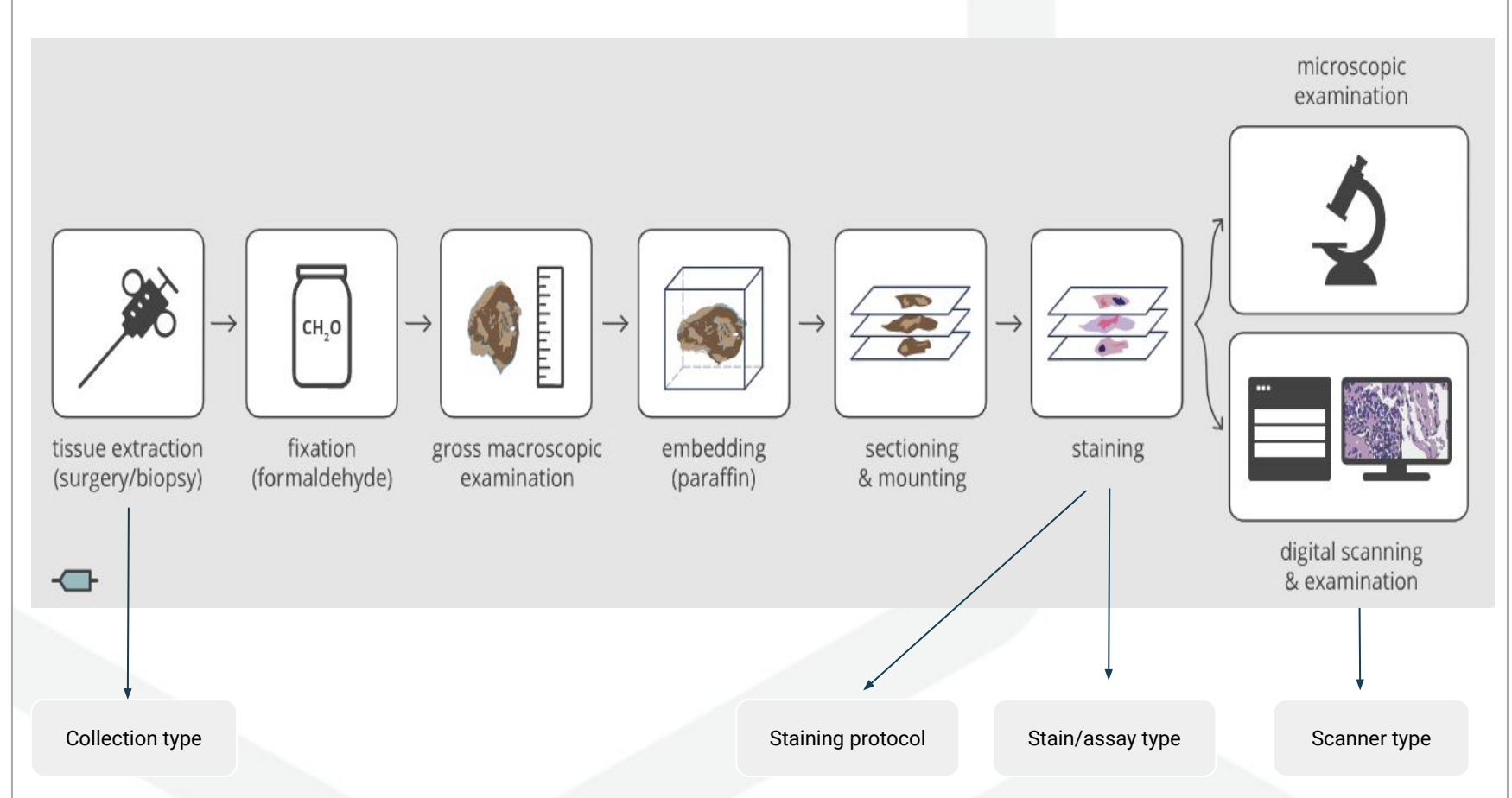


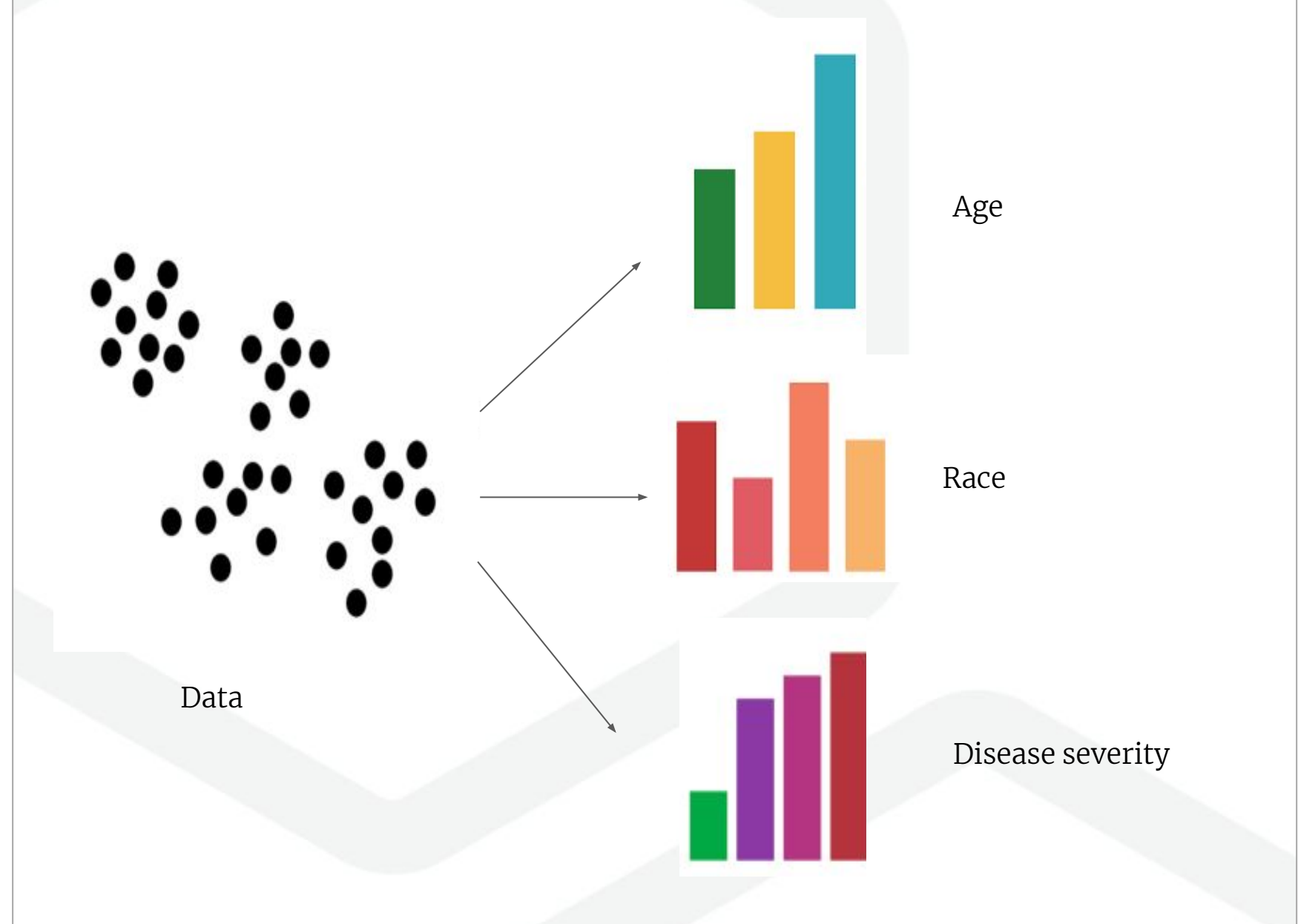
Syed Ashar Javed*, Dinkar Juyal*, Shreya Chakraborty*, Zahil Shanis*, Harsha Pokkalla, Aaditya Prakash
 {firstname.lastname}@pathai.com

Variability in Pathology



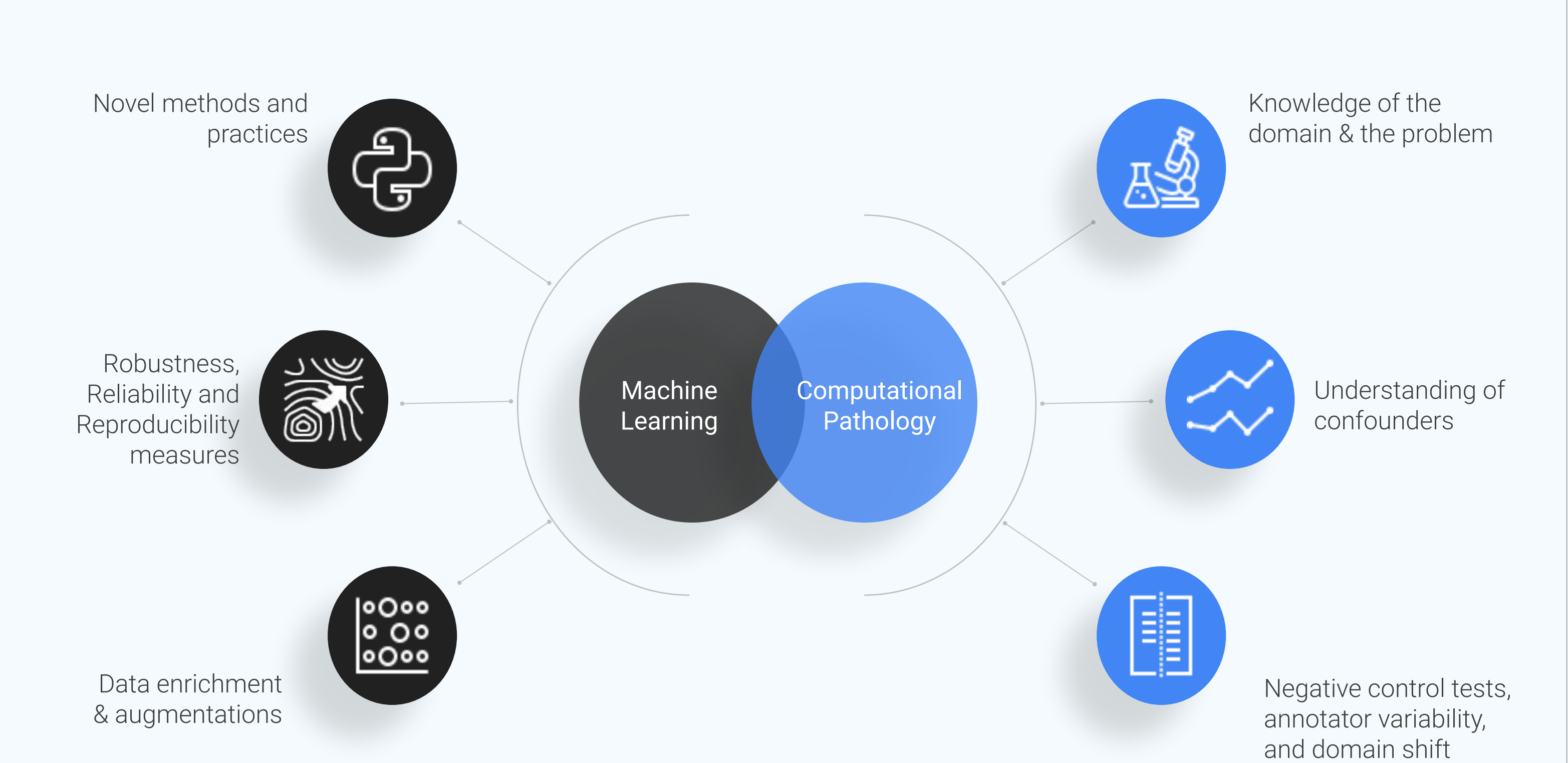
- Pathology images are
 - Gigapixels in size
 - Heterogeneous in terms of sample preparation
 - Susceptible to spurious correlations
 - Expensive and laborious to annotate
 - Require understanding of underlying causal structures

Confounders

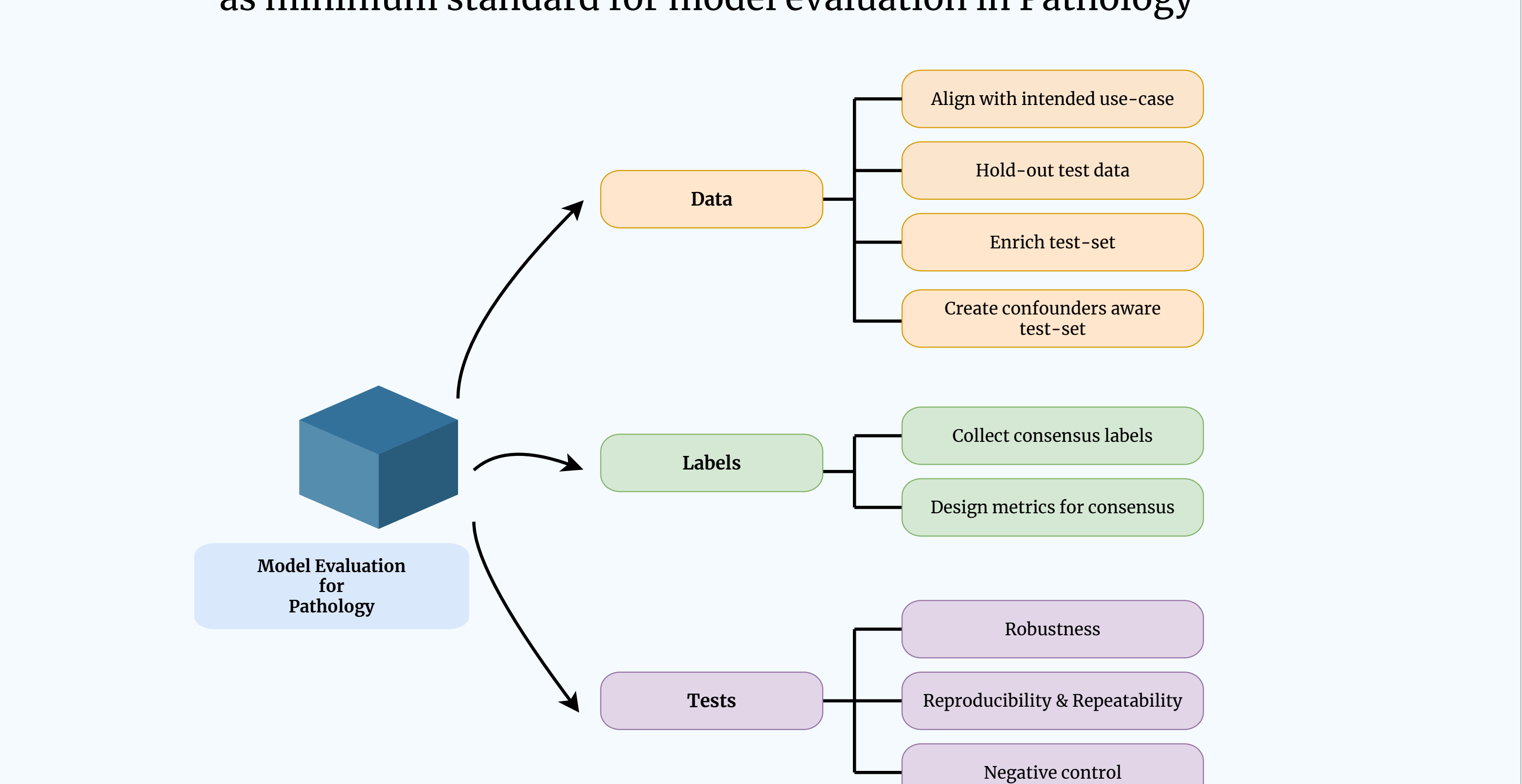


- Enrich held-out test-data for any sub-strata of interest
- Perform correlation test with any metadata that can be a possible confounder
- Design stratified evaluation metrics based on intended use case

Motivation - Model evaluation for pathology is hard!

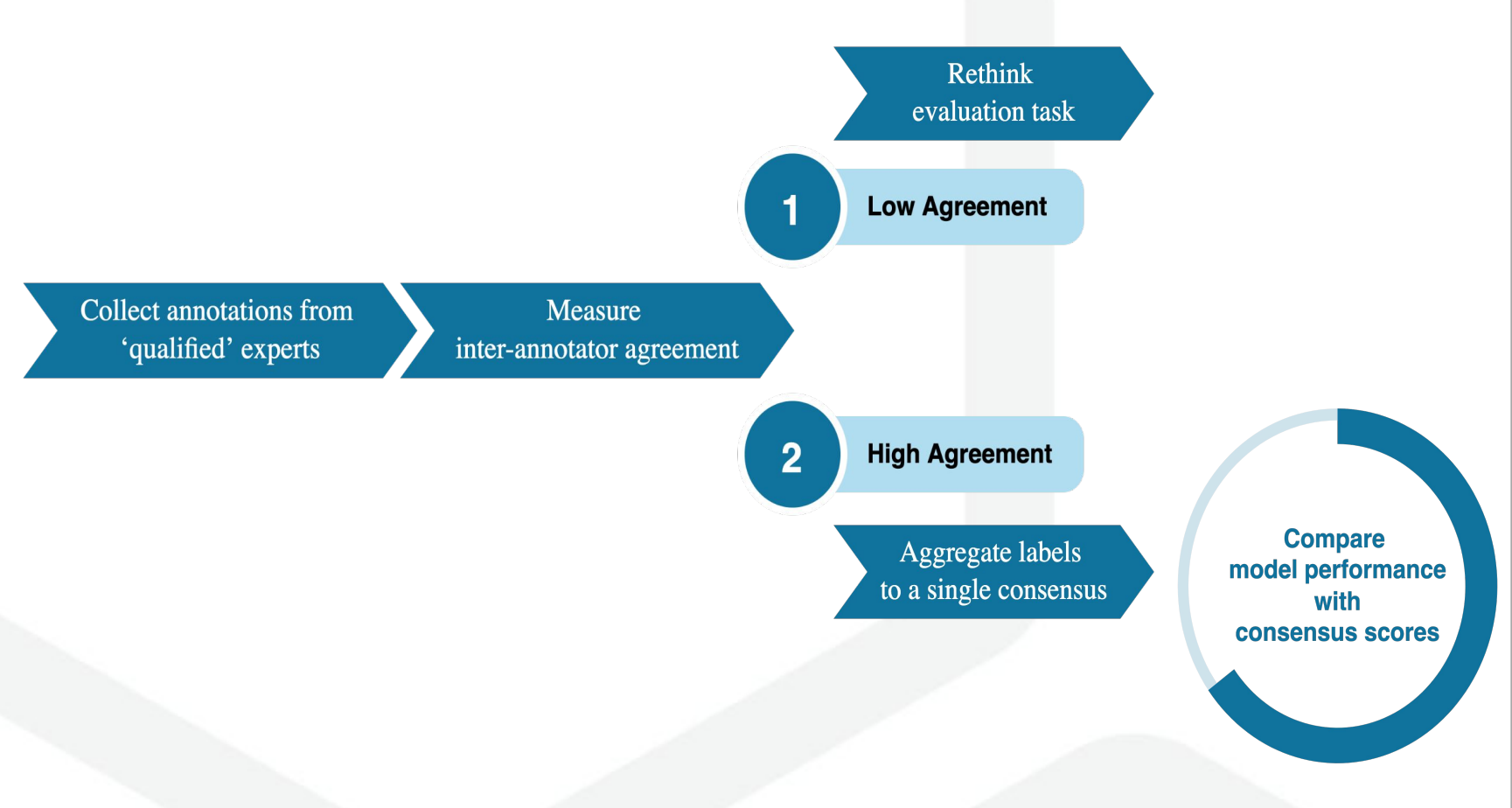


We recommend stratification & enrichment of test data, consensus-aware metrics & robustness to data variations as minimum standard for model evaluation in Pathology



- Data**
 - Align with intended use-case
 - Hold-out test data
 - Enrich test-set
 - Create confounders aware test-set
- Labels**
 - Collect consensus labels
 - Design metrics for consensus
- Tests**
 - Robustness
 - Reproducibility & Repeatability
 - Negative control

Inconsistent Labels



- In pathology, labels are not always reliable
 - Poor label reliability leads to poor model evaluation
- Quantifying subjectivity of labels in both intra and inter-labeller setting is necessary
- Choice of metrics should depend on measuring this reliability well. E.g
 - Kappa Score
 - Intra-class correlation

Tests for Reliability

