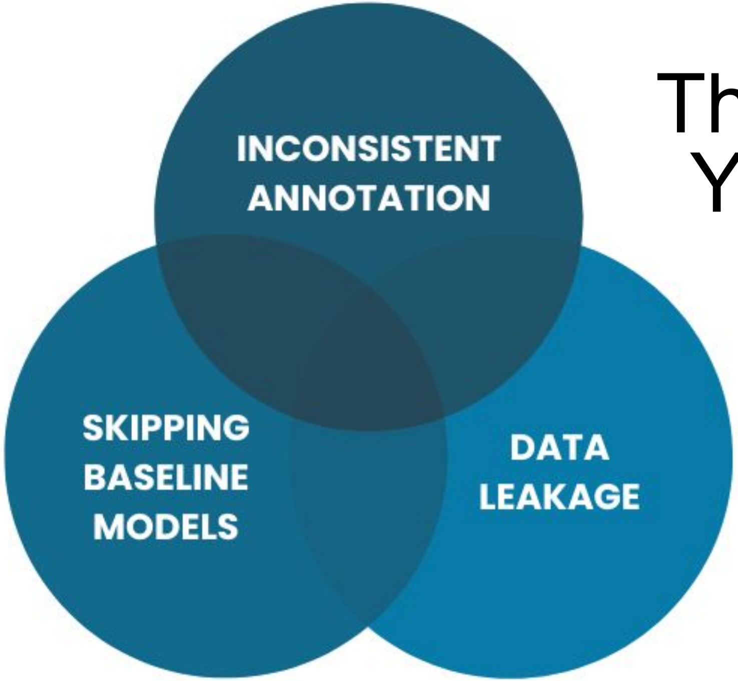


Three Critical Mistakes Derailing Your Computer Vision Projects

Uncover common pitfalls degrading your models

Heather Couture

April 2, 2025
10:30 am EDT

30 minutes + Q&A

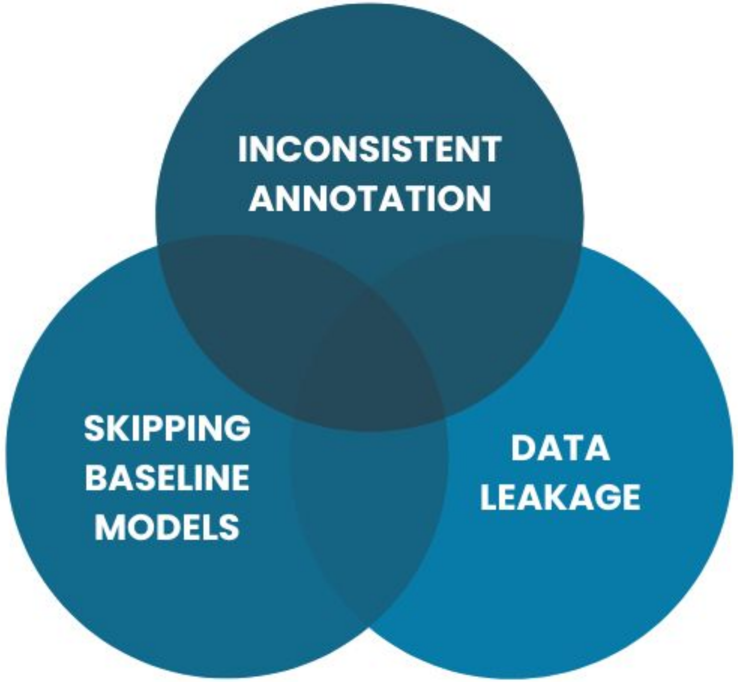
Pixel Scientia LABS

Goals of this Webinar

How these problems manifest

Why they’re important to address

Best practices for handling each



Who am I?

- Heather Couture
- PhD in Computer Science from University of North Carolina
- Contributor to Scientific American, The Pathologist, IEEE Spectrum
- Newsletter and podcast

Computer Vision Insights
by Pixel Scientia Labs

IMPACT AI
WITH HEATHER COUTURE

- Computer vision consultant

OWKIN DECIPHEX ULTIVUE WattTime Leica MICROSYSTEMS QRITIVE

Critical Mistake #1: Inconsistent Annotation Processes

Increased noise in labels

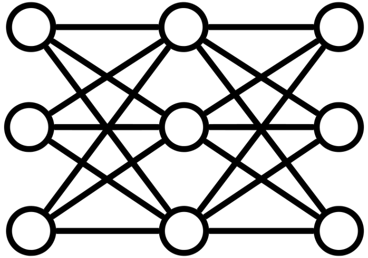
Reduced accuracy

Biased predictions

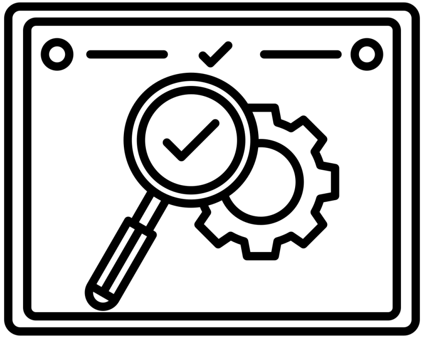
Potential harm in critical applications

Delayed model development

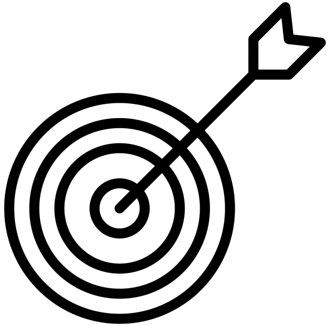
The Role of Annotation



Training models

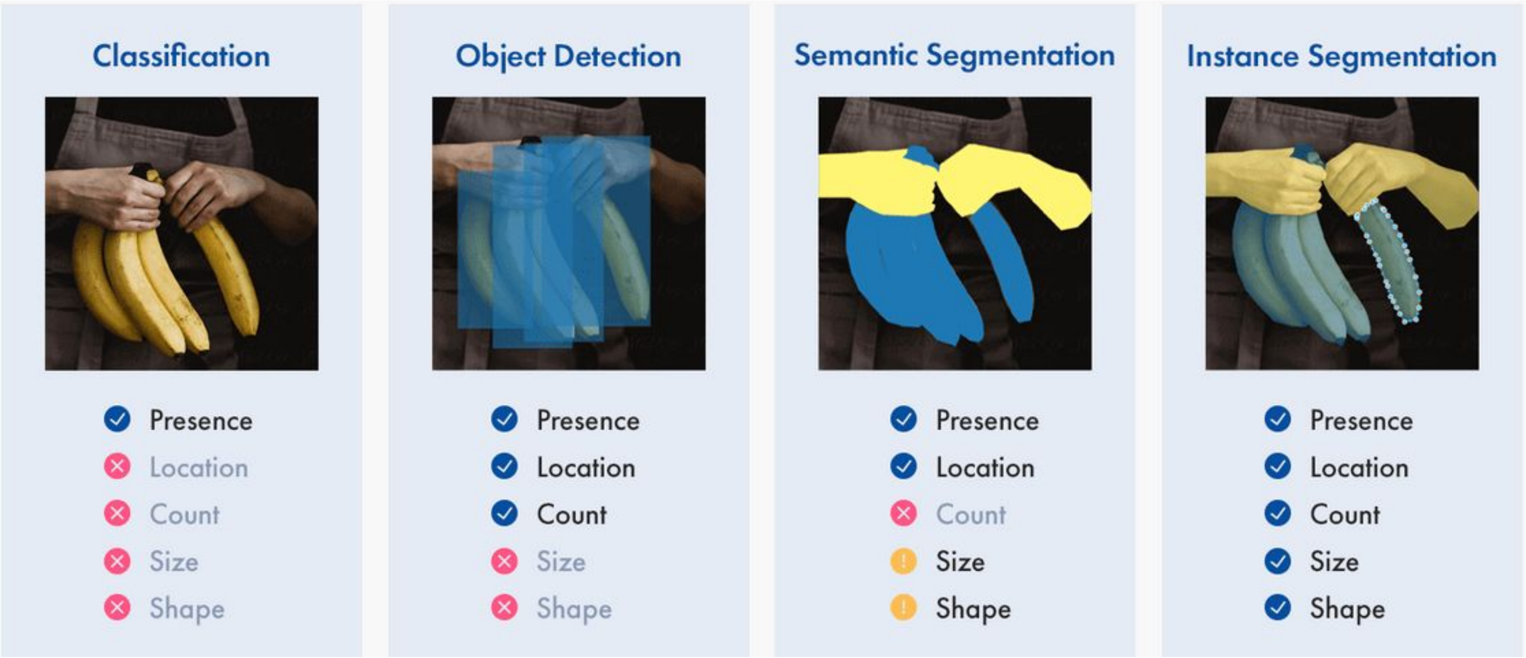


Validating model performance



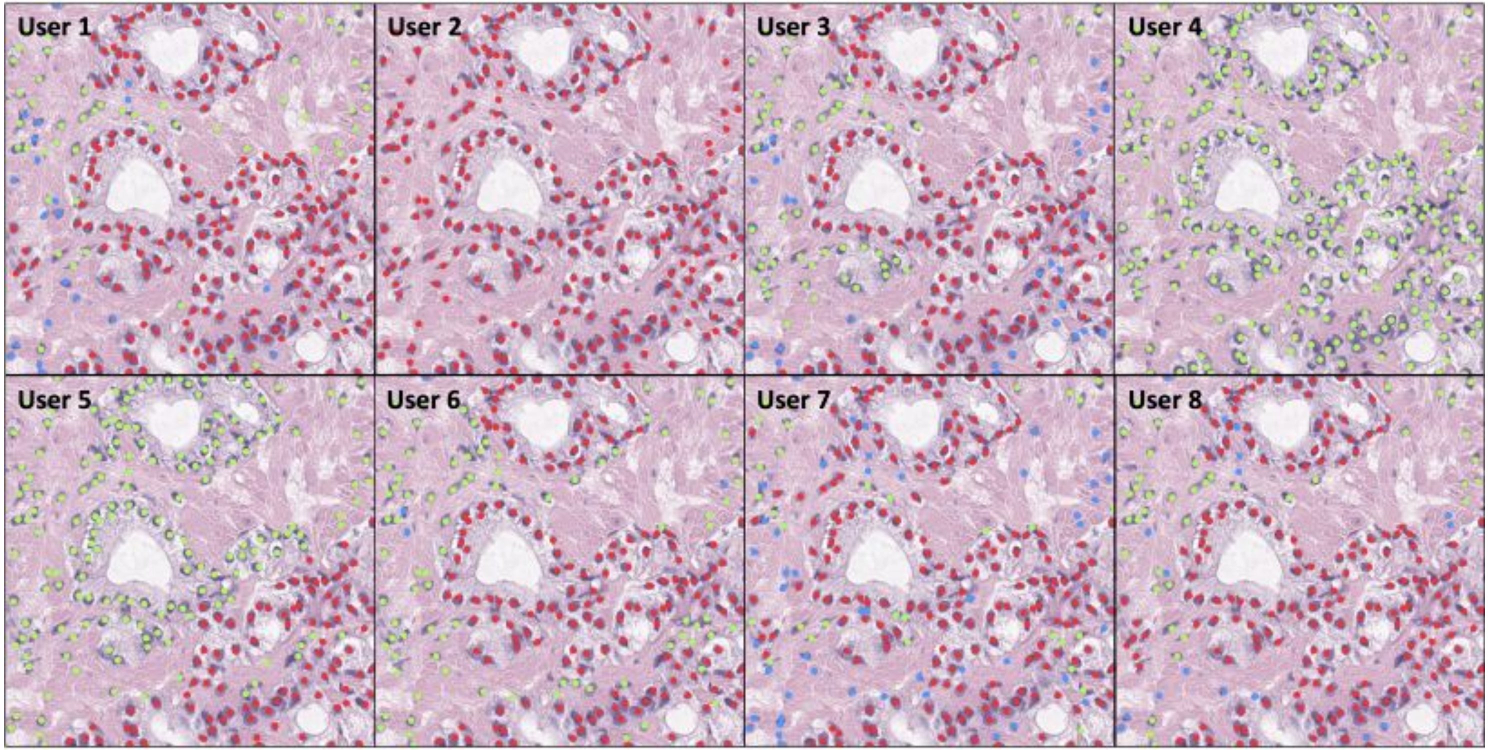
Improving model accuracy

Common Types of Annotation



Source: Image Annotation for Computer Vision, https://www.cloudfactory.com/image-annotation-guide

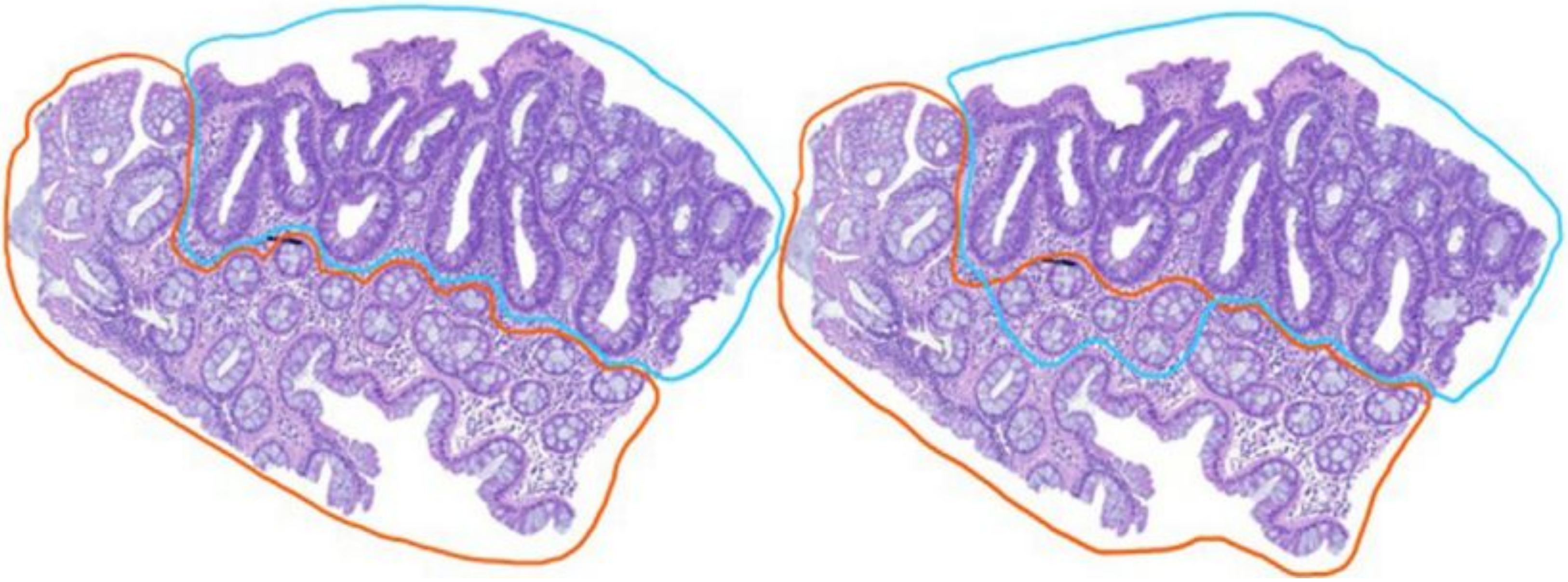
Example: Object Classification Discrepancies



Red: tumor cell
Blue: lymphocyte
Green: other cell

Source: Kang, Variability Matters: Evaluating inter-rater variability in histopathology for robust cell detection, 2022

Example: Boundary Inconsistencies



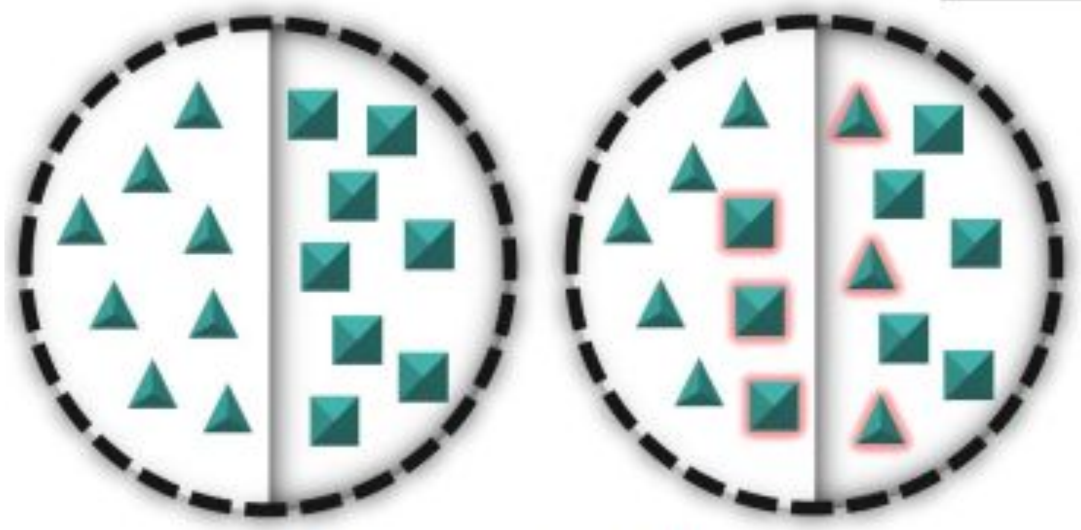
Source: Montezuma, Annotating for Artificial Intelligence Applications in Digital Pathology: A Practical Guide for Pathologists and Researchers, 2022

Example: Incomplete Annotations

Source: https://pixabay.com/photos/oranges-fruits-grove-orange-trees-1117628/

Inconsistent Annotation Causes a Posterior Shift

$P(Y_S|X_S) \neq P(Y_T|X_T)$



S = Source
T = Target
X = features
Y = labels

Measuring Inter-rater Agreement

Percentage Agreement

n_{concur} / n

Cohen's Kappa

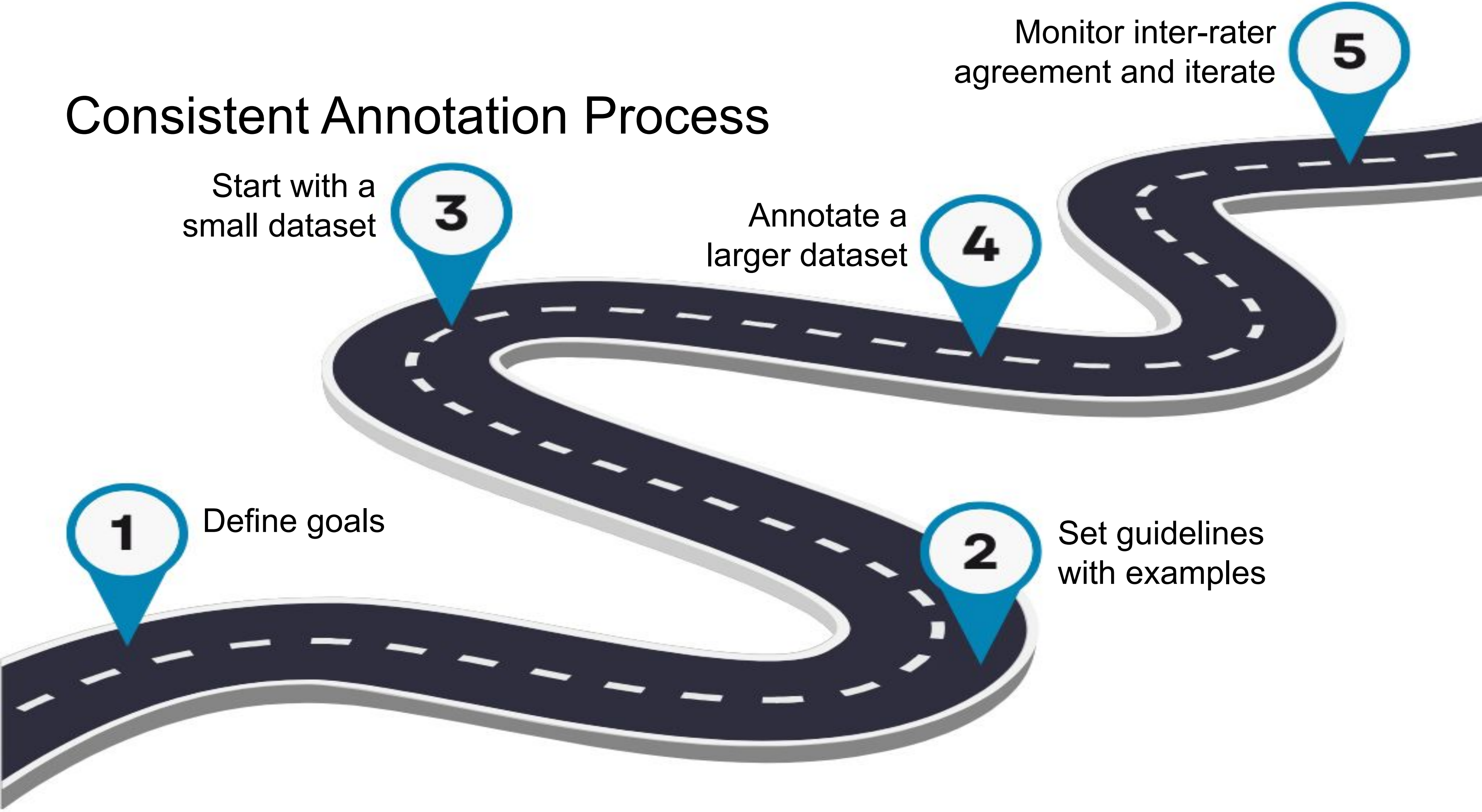
$K = (p_o - p_e) / (1 - p_e)$

p_o: observed agreement proportion

p_e: expected agreement by chance

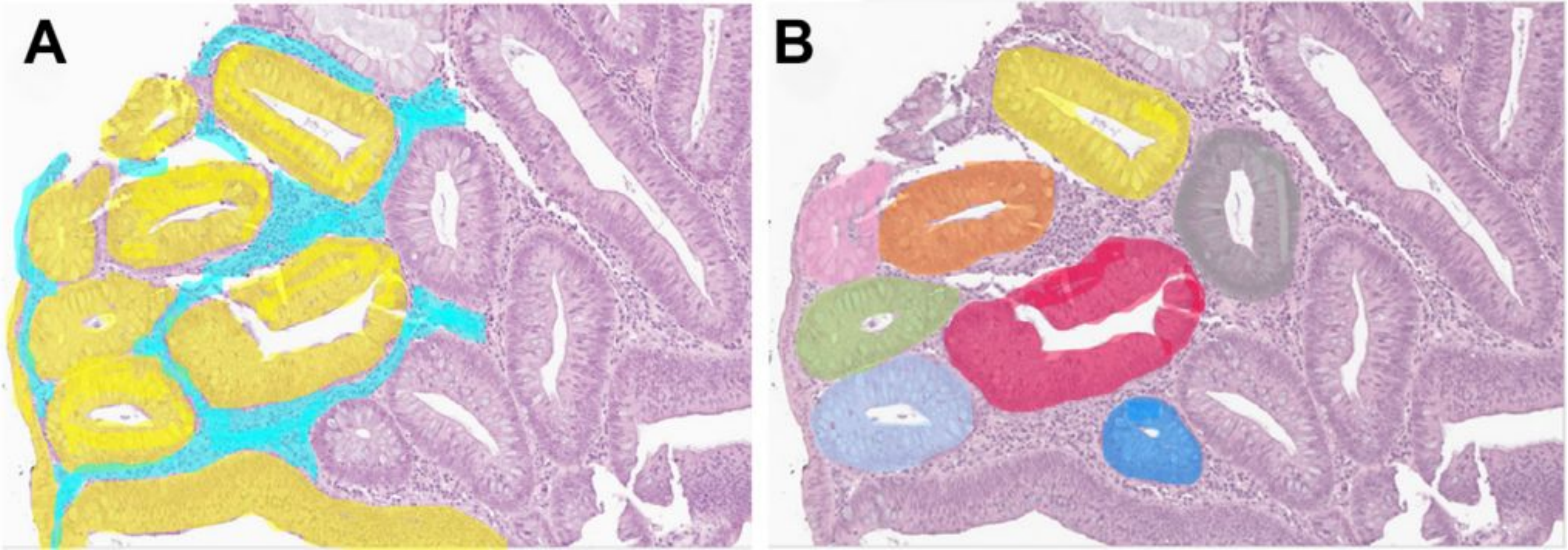
	Labels	Percentage agreement	Cohen's Kappa
Annotator 1	1 1 1 0	3/4 = 0.75	$p_o = 0.75$ $p_e = 0.5$ $K = 0.5$
Annotator 2	0 1 1 0		

Consistent Annotation Process



1) Define Goals

Semantic vs. Instance Segmentation



Source: Montezuma, Annotating for Artificial Intelligence Applications in Digital Pathology: A Practical Guide for Pathologists and Researchers, 2022

Is Manual Annotation Needed?

Source: Couture, Towards Tracking the Emissions of Every Power Plant on the Planet, NeurIPS 2020

2) Set Guidelines with Examples

Examples of accurate and inaccurate annotations

Detailed descriptions of each example

Add new cases when needed

Use simple, straightforward language

Avoid complex or subjective criteria

3) Start with a Small Dataset

Representative but manageable in size

Annotate following guidelines

Gather feedback from annotators

Discuss examples with disagreement

Update guidelines

Repeat if needed

4) Annotate a Larger Dataset

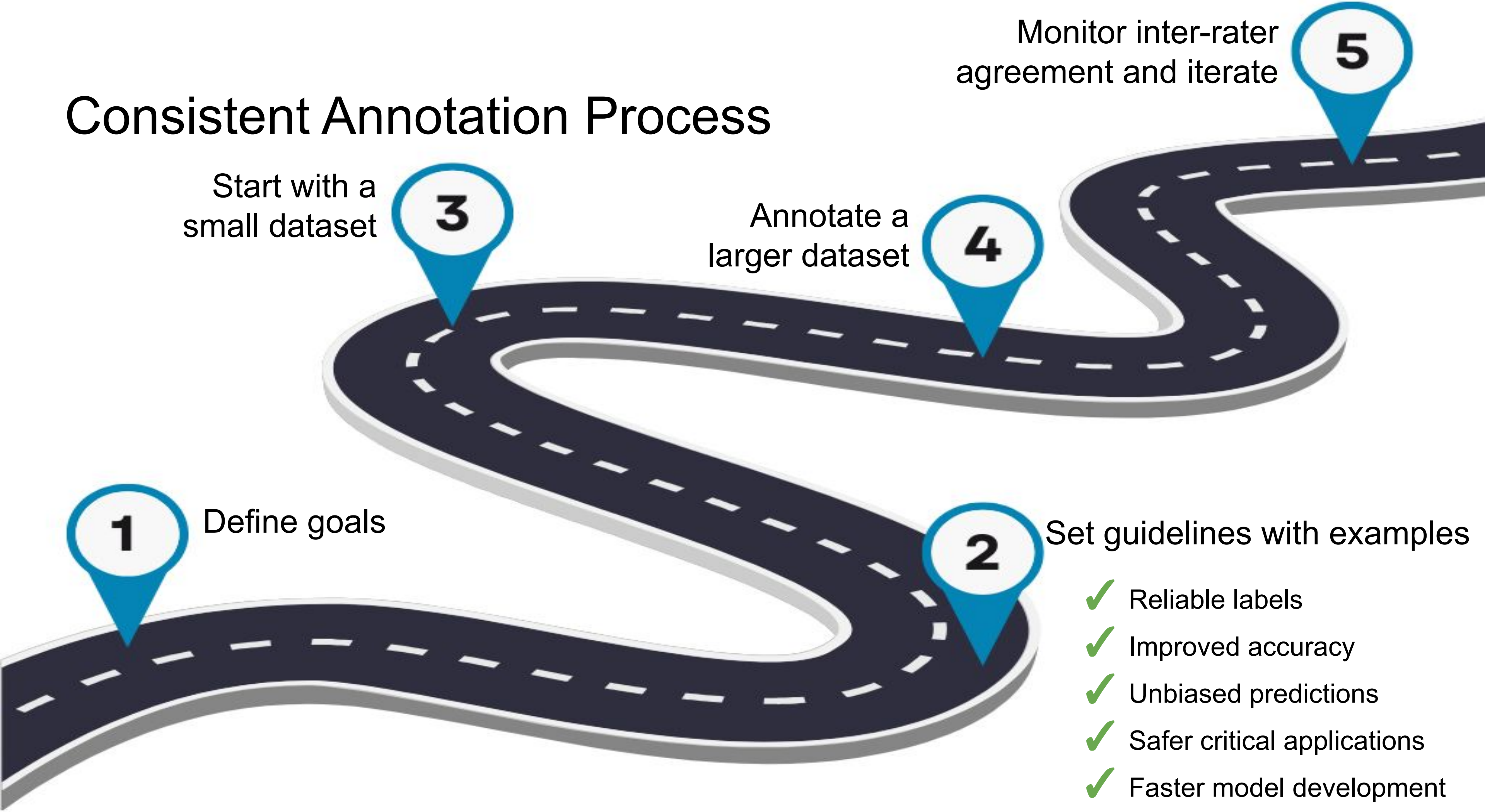
Train annotators using the refined guidelines and developed examples

Break into batches and measure inter-rater agreement after each round

If agreement is low:

- Review the guidelines and examples
- Identify areas of ambiguity or confusion and clarify the criteria
- Provide additional feedback and training to annotators
- Revise guidelines

Consistent Annotation Process



Critical Mistake #2: Skipping Baseline Models

Harder to pinpoint data problems

Difficult to tell whether complex model is worth it

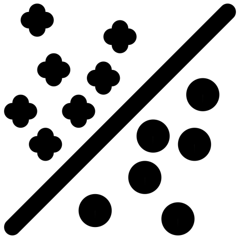
Inability to gauge progress

Wasted time and resources

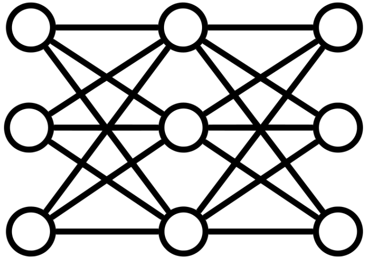
Less informed, less efficient, less successful

Benefits: Establish a Performance Benchmark

"Does my complex model outperform a simpler one?"



vs.



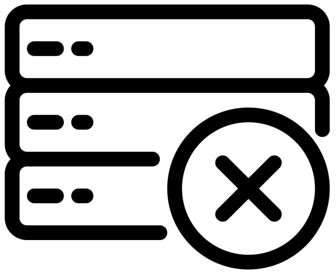
"Is the extra complexity worth it?"

Benefits: Understanding the Dataset

Low signal strength

Difficult-to-classify classes or observations

Annotation inconsistencies

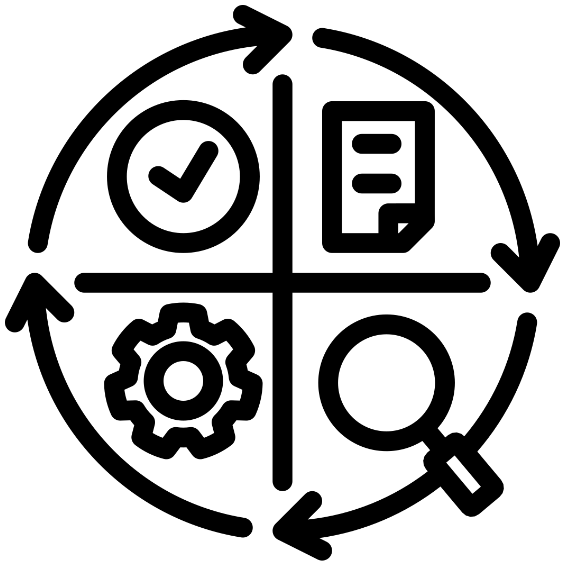


Benefits: Facilitating Faster Iteration

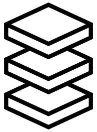
Quick and easy to build

Rapid experimentation and iteration

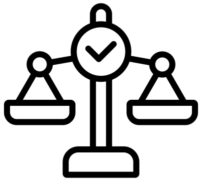
Gain new insights quickly



Baseline Best Practices: Basic Data Processing



Small dataset

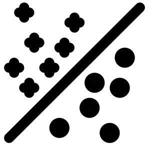


Balanced dataset

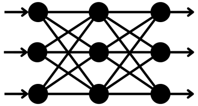
Minimal augmentation

Model with other variables like demographics

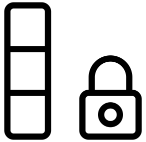
Baseline Best Practices: Simple and Fast Model



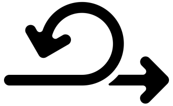
Basic architectures



Pre-trained models



Freeze pre-trained weights



Fewer epochs

Start with a Simple Baseline Model

Easier to pinpoint data problems

Justify model complexity

Measure progress

Faster iteration

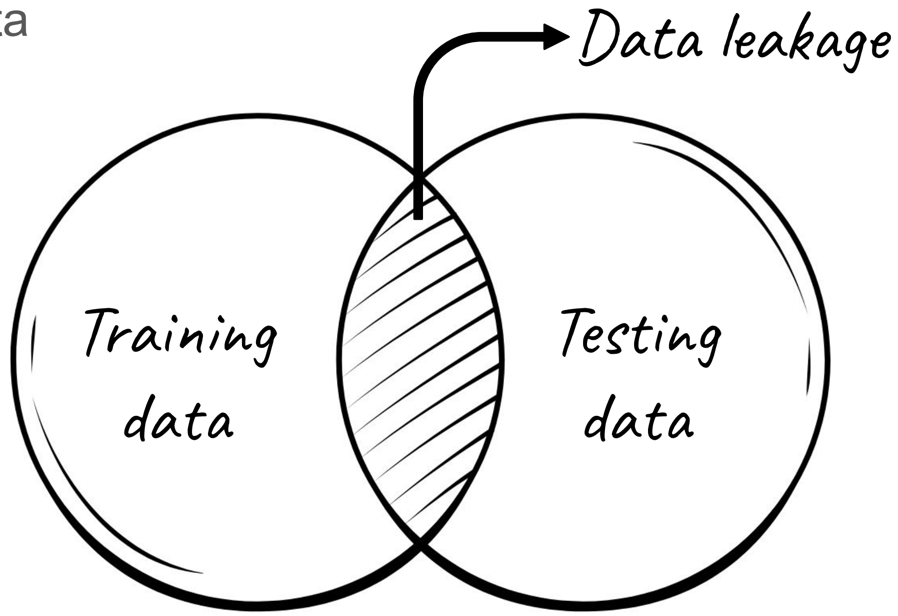
Critical Mistake #3: Data Leakage

Overfitting

Lower ability to generalize to unseen data

Inflated model performance

Misguided decisions



Causes of Data Leakage

Train-test overlap

Group leakage

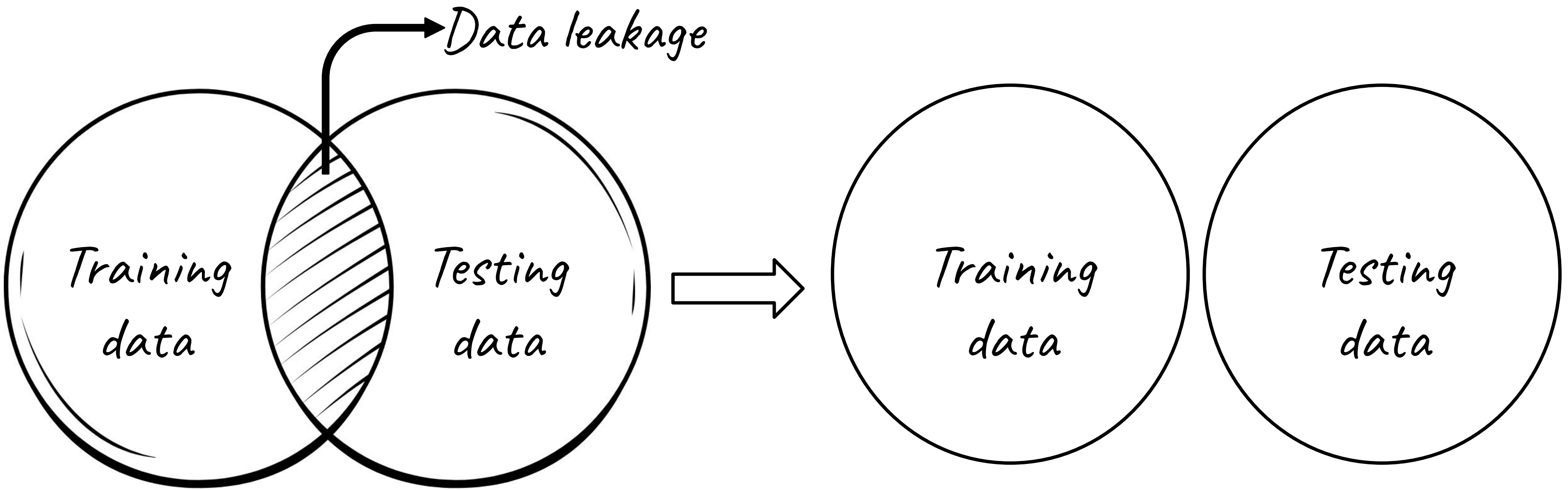
Target leakage

Feature leakage

Data transformation leakage

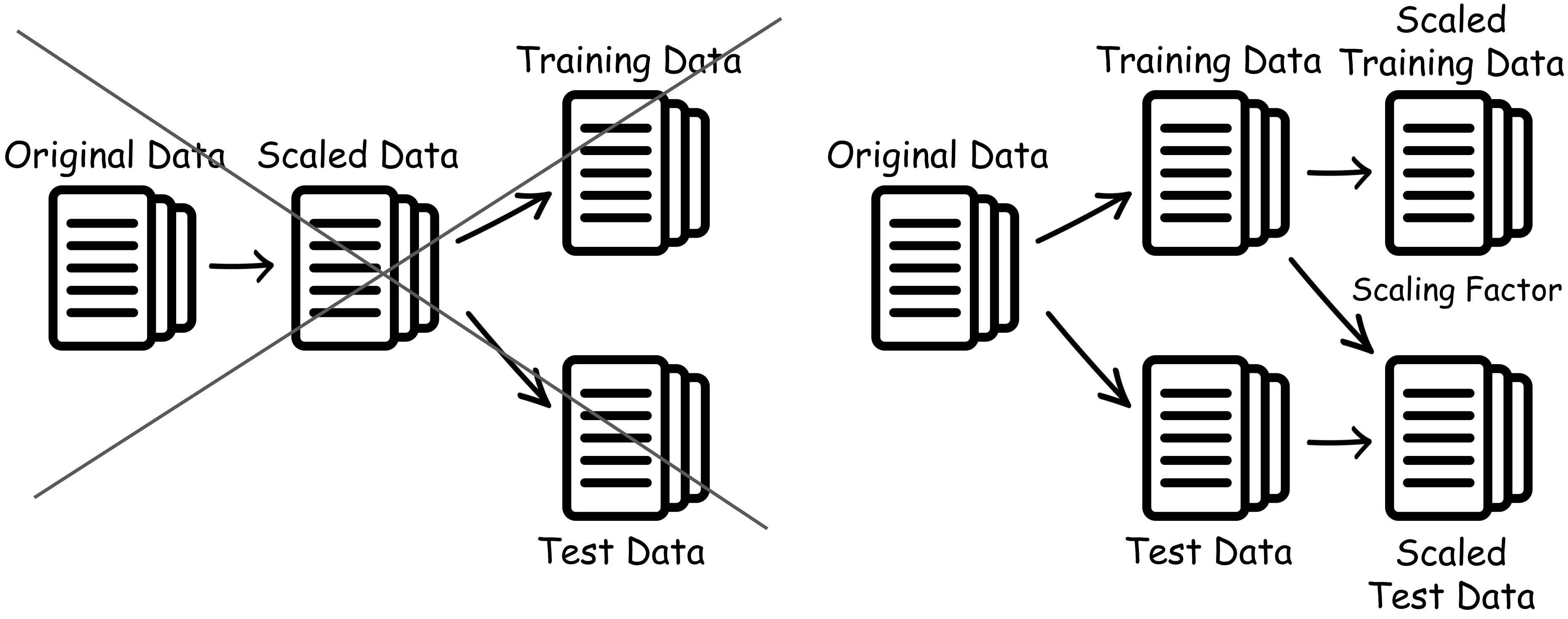
Validation leakage

Prevention: Proper Train-Test Split

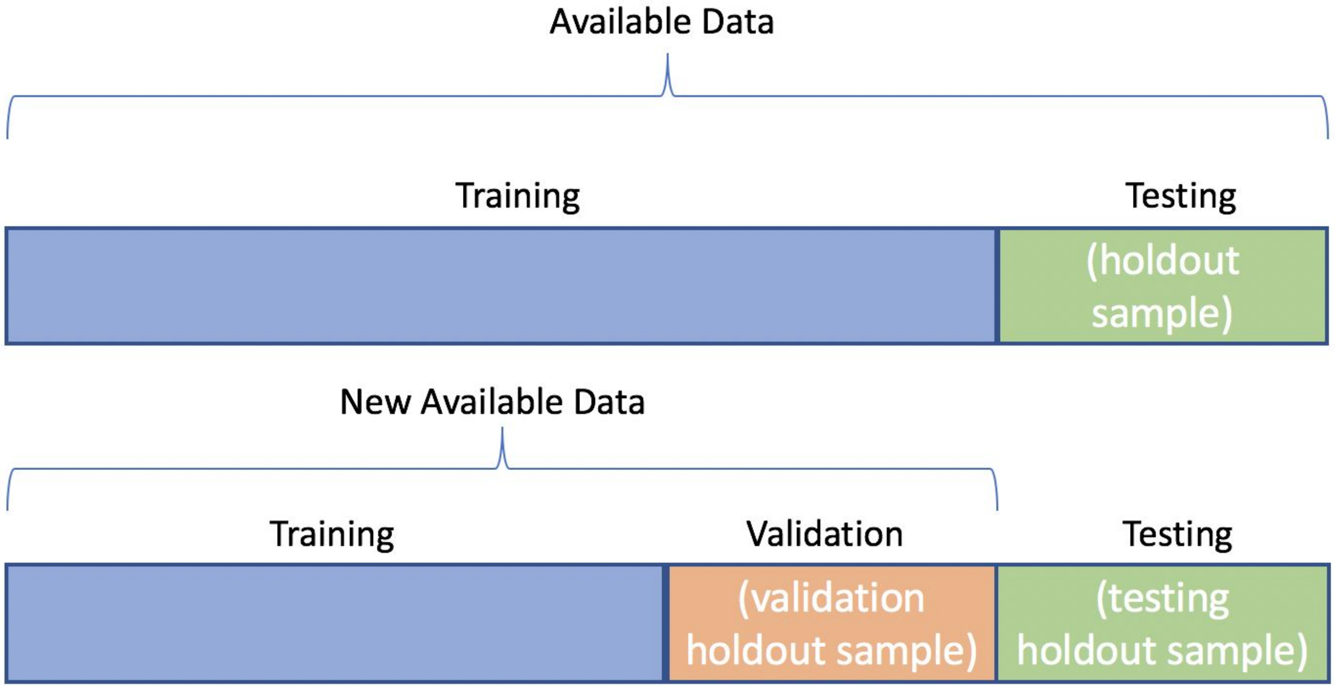


No overlap in patients,
even if a different lesion or part of the body,
regardless of how small your dataset is

Prevention: Transformation and Feature Selection After Split



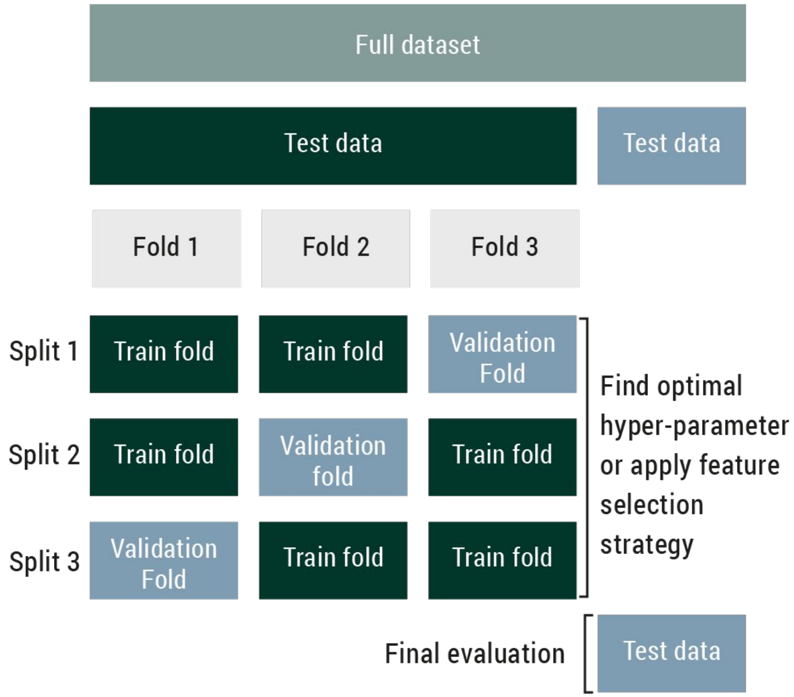
Prevention: Separate Validation Set for Tuning Hyperparameters



Source: https://algotrading101.com/learn/train-test-split/

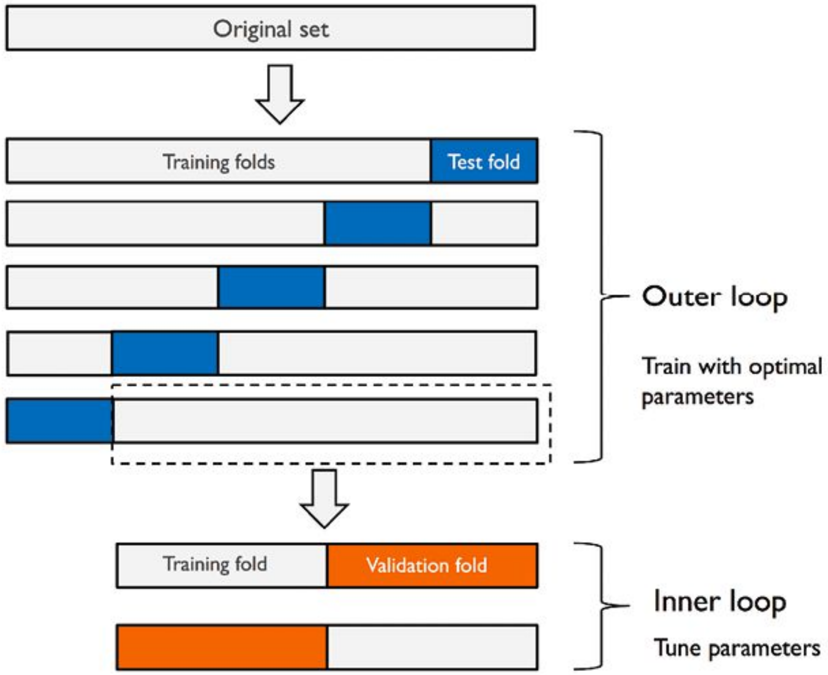
Prevention: Careful Cross-Validation

Simple Cross-Validation



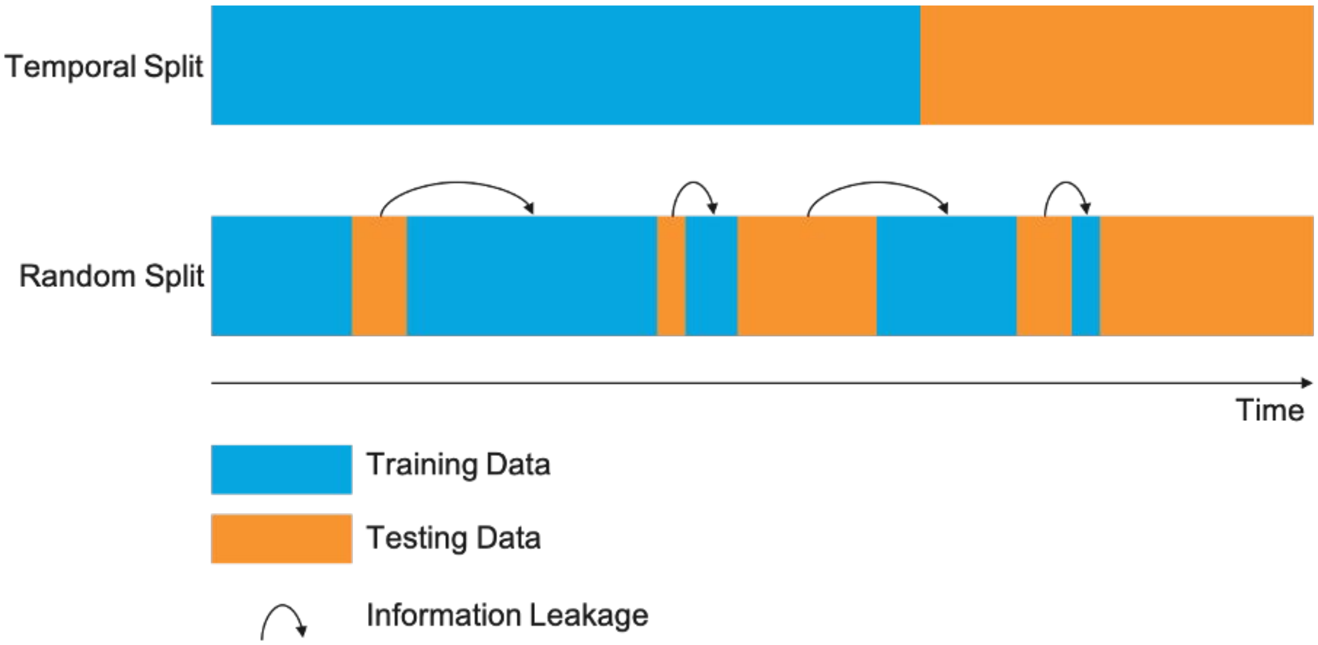
Source: https://datasciencetalent.co.uk/tam-tran-shares-insights-on-how-to-fix-common-cross-validation-pitfalls/

Nested Cross-Validation



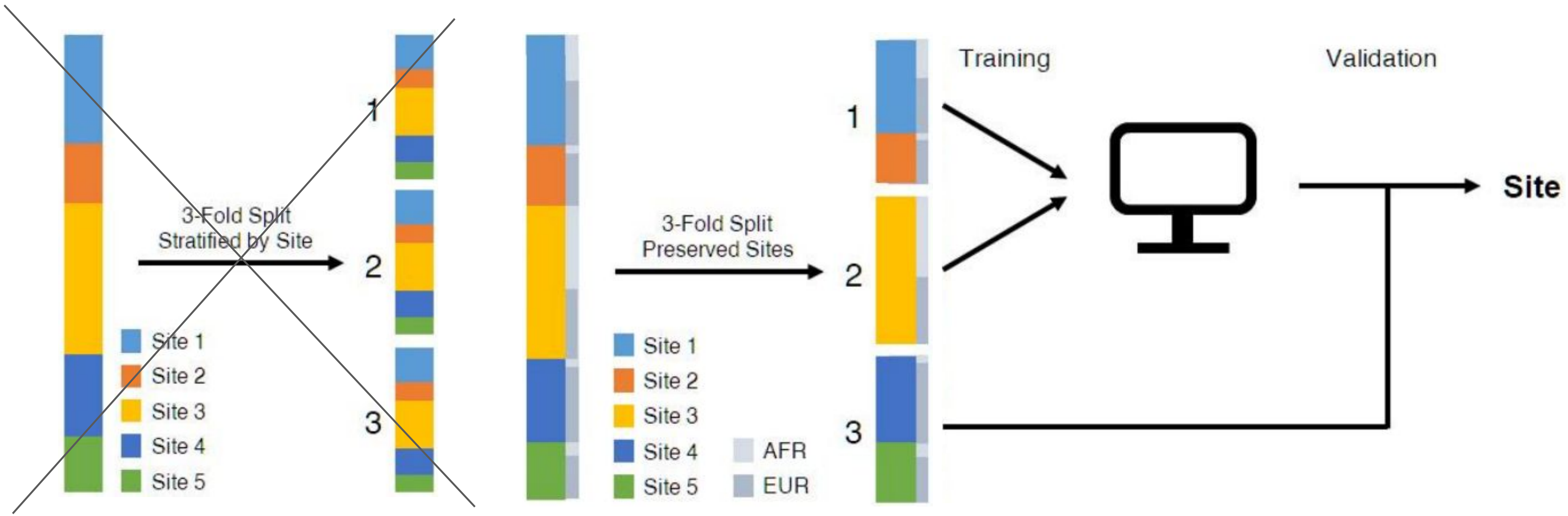
Source: https://vitalflux.com/python-nested-cross-validation-algorithm-selection/

Prevention: Temporal Split



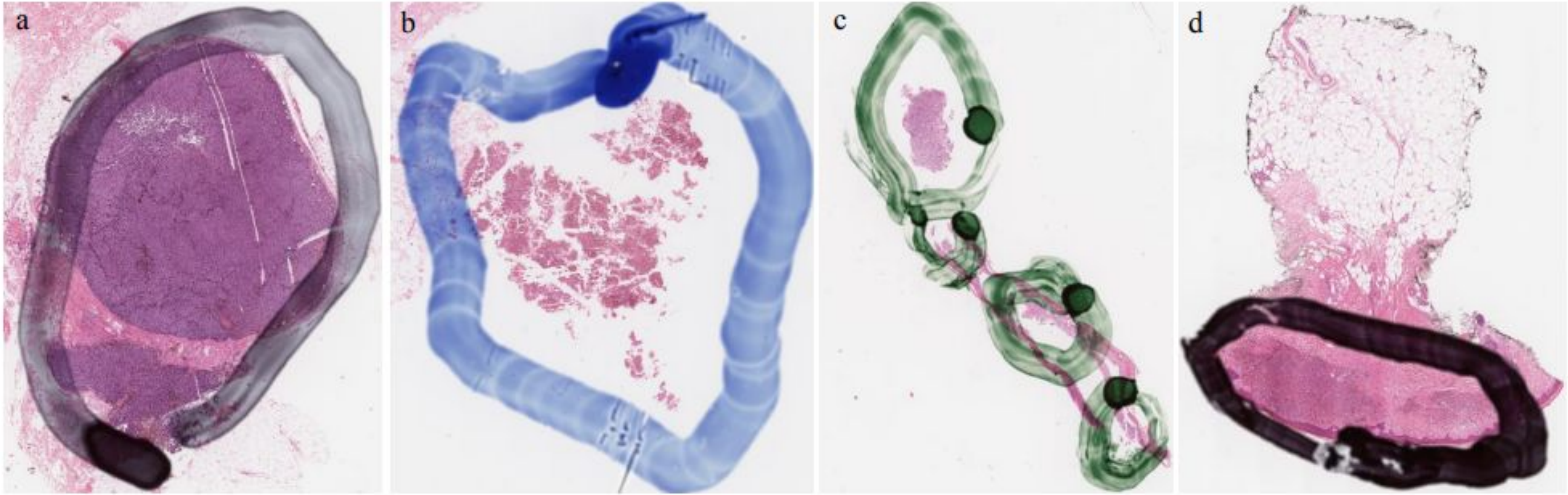
Source: https://c3.ai/glossary/data-science/information-leakage/

Prevention: Group Split



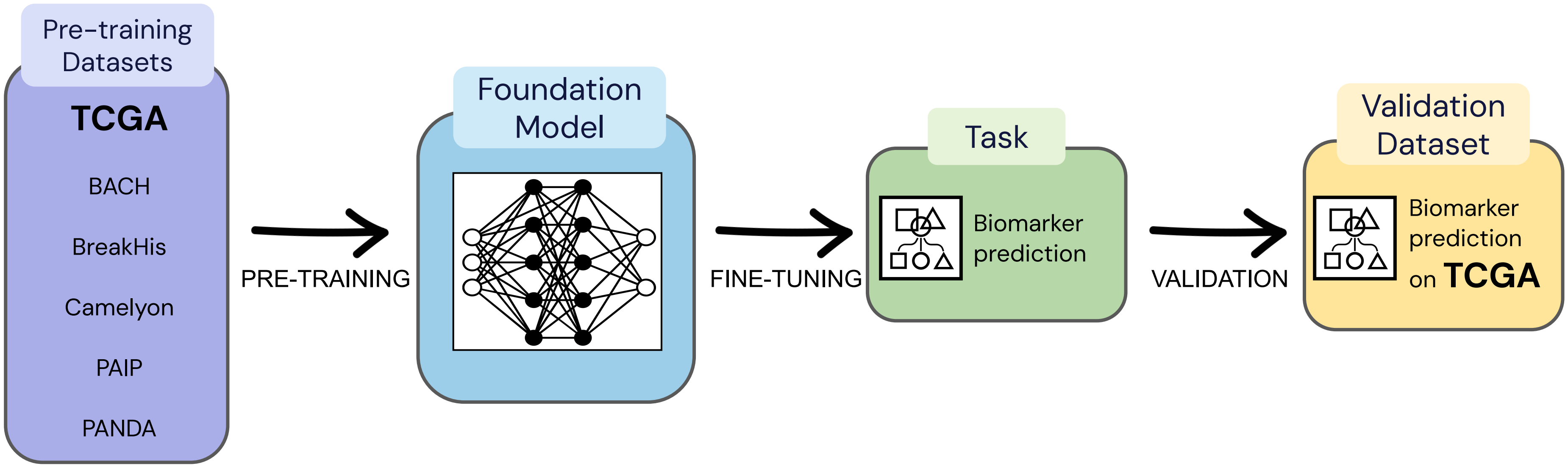
Source: Howard, The Impact of Digital Histopathology Batch Effect on Deep Learning Model Accuracy and Bias, 2020

Prevention: Quality Control



Source: Venkatesh, Restoration of Marker Occluded Hematoxylin and Eosin Stained Whole Slide Histology Images Using Generative Adversarial Networks, 2019

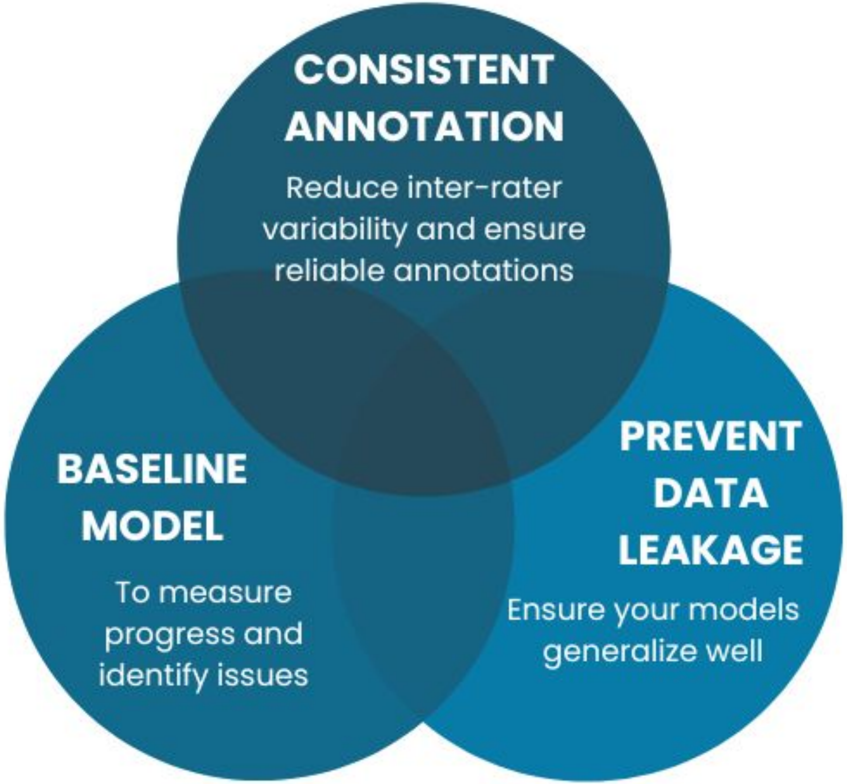
Prevention: Be Aware of Pre-training Datasets



Prevent Data Leakage

1) Proper train-test split
2) Transformation and feature selection after split
3) Careful cross-validation
4) Quality control
5) Be aware of pre-training datasets

Key Takeaways



Are you caught in one of these pitfalls and unsure of the best path forward?

Who I work with:

- Founders and other leaders
- Their technical team

Example results:

- A roadmap to streamline model development
- Break through roadblocks
- Keep up with AI trends and innovations
- Boost investor confidence

Advisory services:

- Monthly strategy call
- Weekly office hours
- Private Slack channel
- And more

Resources

Team workshops: https://pixelscientia.com/workshops

Current topics:

Mastering Distribution Shift in Computer Vision

Harnessing the Power of Foundation Models for Pathology

Other consulting services: heather@pixelscientia.com

Pixel Scientia LABS