OCTOBER 2025

Oyster MSX Pipeline

A Human-in-the-Loop System for Accelerating MSX Diagnosis in Oyster Histology

BY JONATHAN E. ZUL LUNA

UNIVERSITY of Prince Edward ISLAND

Science Atlantic

Mentors:
Dr. Kristina Kupferschmidt, Dr. Andrew Godbout, Dr. David Groman

The Economic Engine

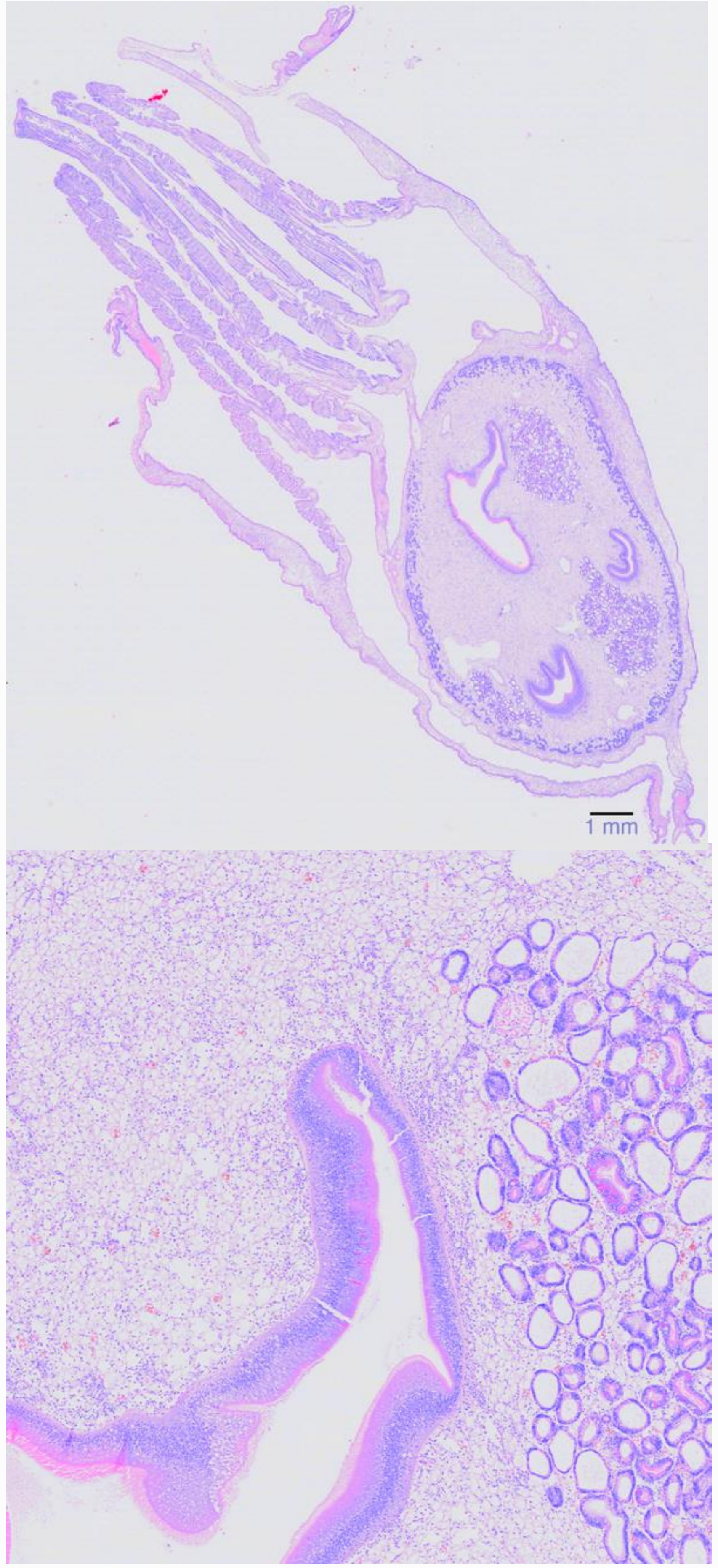
PEI’s Oyster Aquaculture Industry

- A Key economic driver for the province, valued at over **$24 million** annually.
- Supports hundreds of jobs in rural and coastal communities.

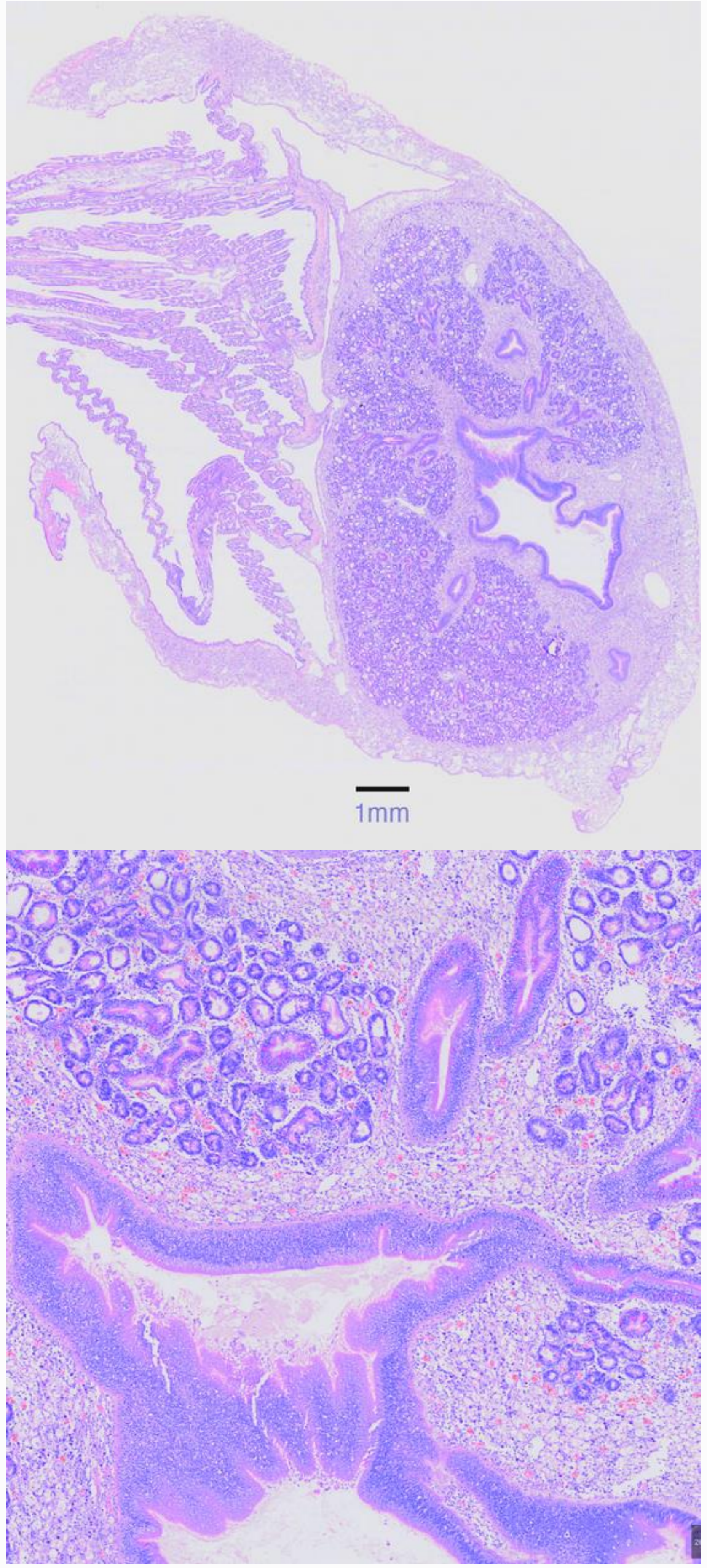
A Microscopic Threat

The MSX Parasite (*Haplosporidium Nelsoni*)

- A protozoan parasite that causes Multinucleated Sphere X (MSX) disease.
- Leads to severe tissue damage and high mortality rates in infected oysters.
- Early and accurate detection is the only way to manage its spread and protect the industry.



Healthy Oyster Tissue

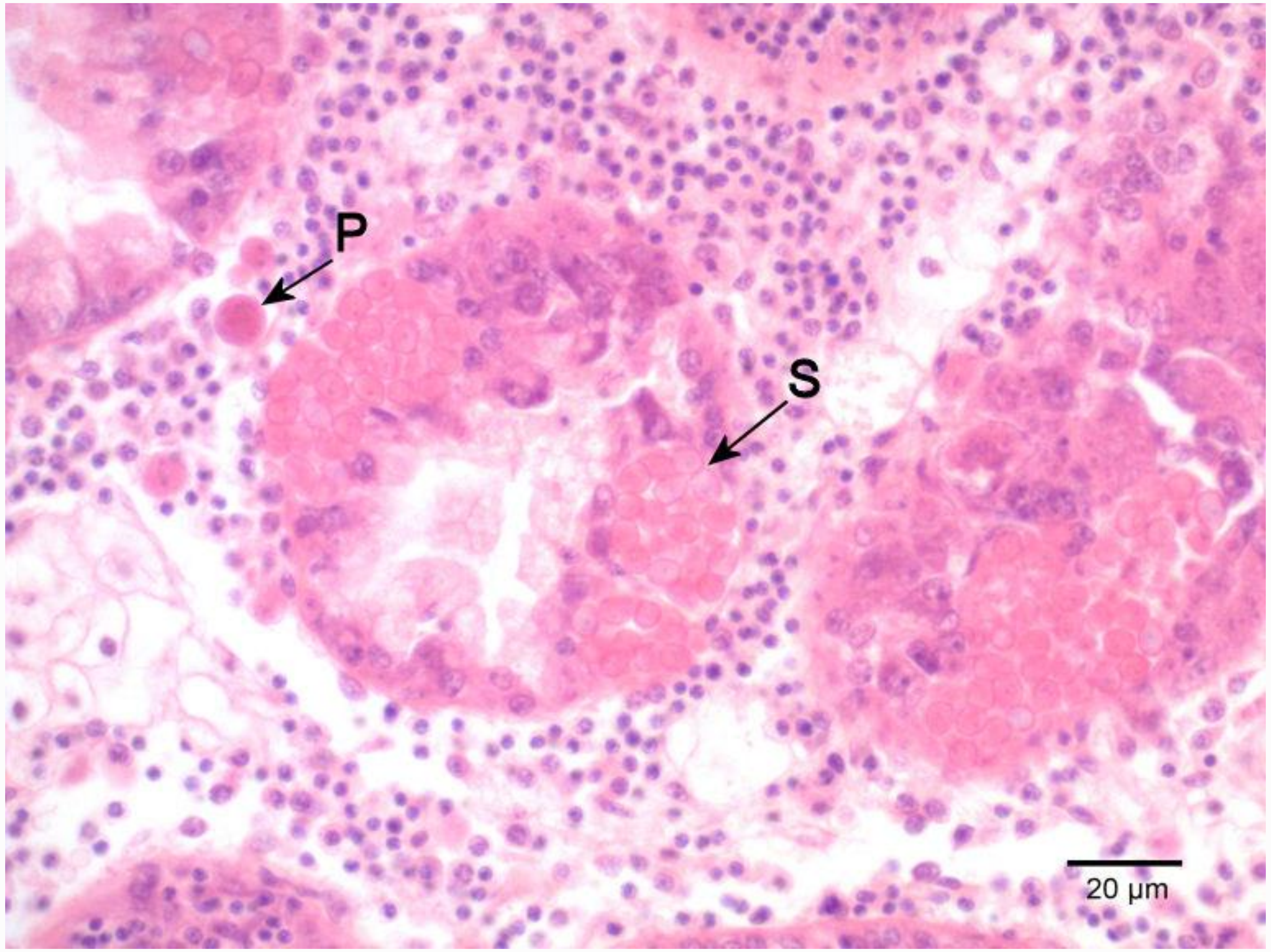


MSX-Infected Tissue

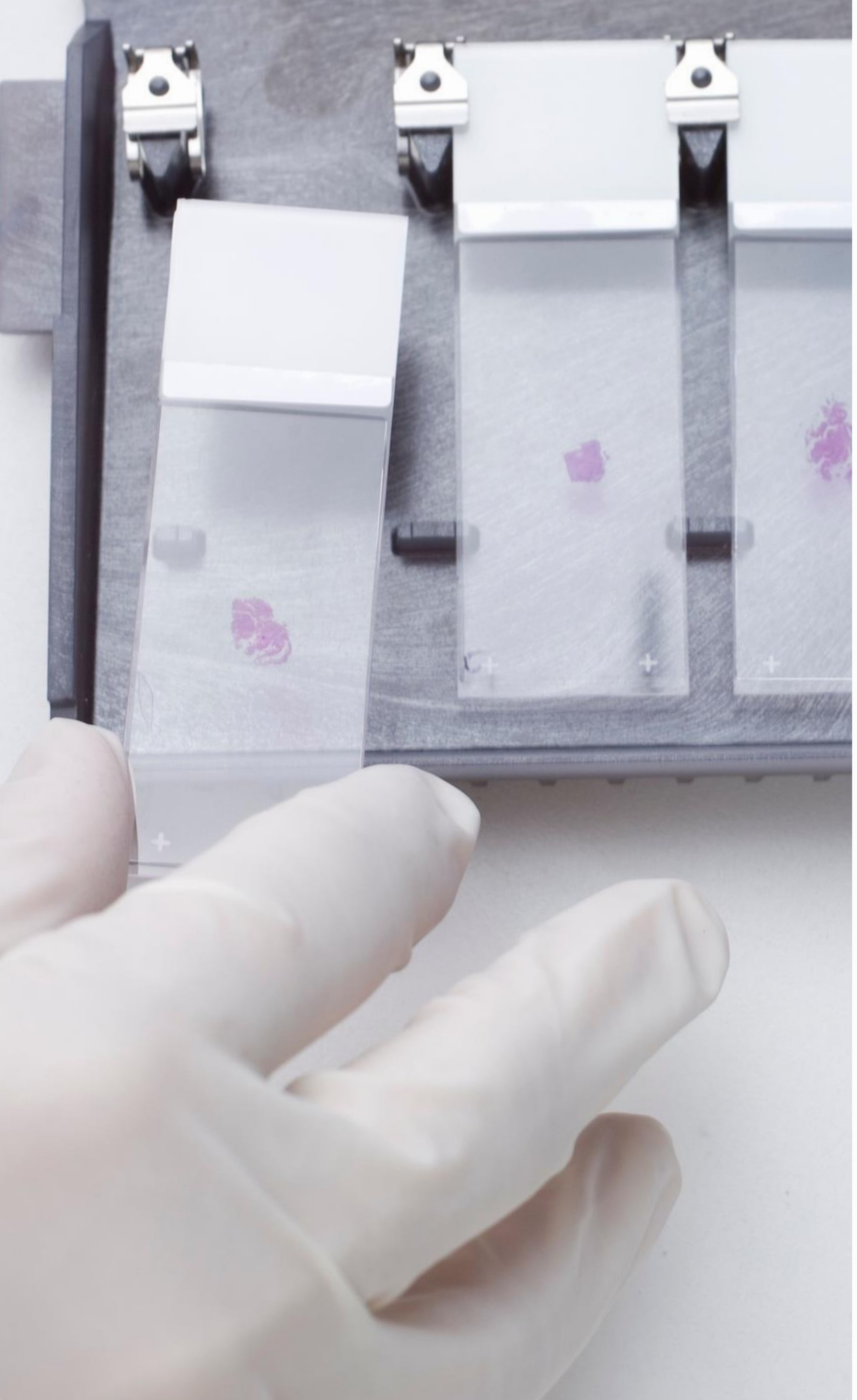
A Microscopic Threat

The MSX Parasite (*Haplosporidium Nelsoni*)

- A protozoan parasite that causes Multinucleated Sphere X (MSX) disease.
- Leads to severe tissue damage and high mortality rates in infected oysters.
- Early and accurate detection is the only way to manage its spread and protect the industry.



MSX-infected oyster tissue, showing the inflammatory response caused by the disease.



The Diagnostic Front Line

The Role of the Atlantic Veterinary College (AVC)

- The AVC is the primary centre for histology-based diagnosis in Atlantic Canada.
- Histology is the gold standard for assessing disease severity, unlike other methods (e.g., PCR).
- Diagnosis relies on manual slide review by a small number of trained pathologists.
- This process is time-consuming, costly, and creates a significant workflow bottleneck.

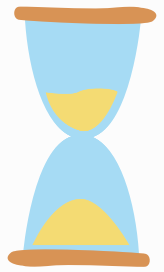
The Histopathology Workflow

The Current Manual Workflow

Oyster Collection

Slide Preparation

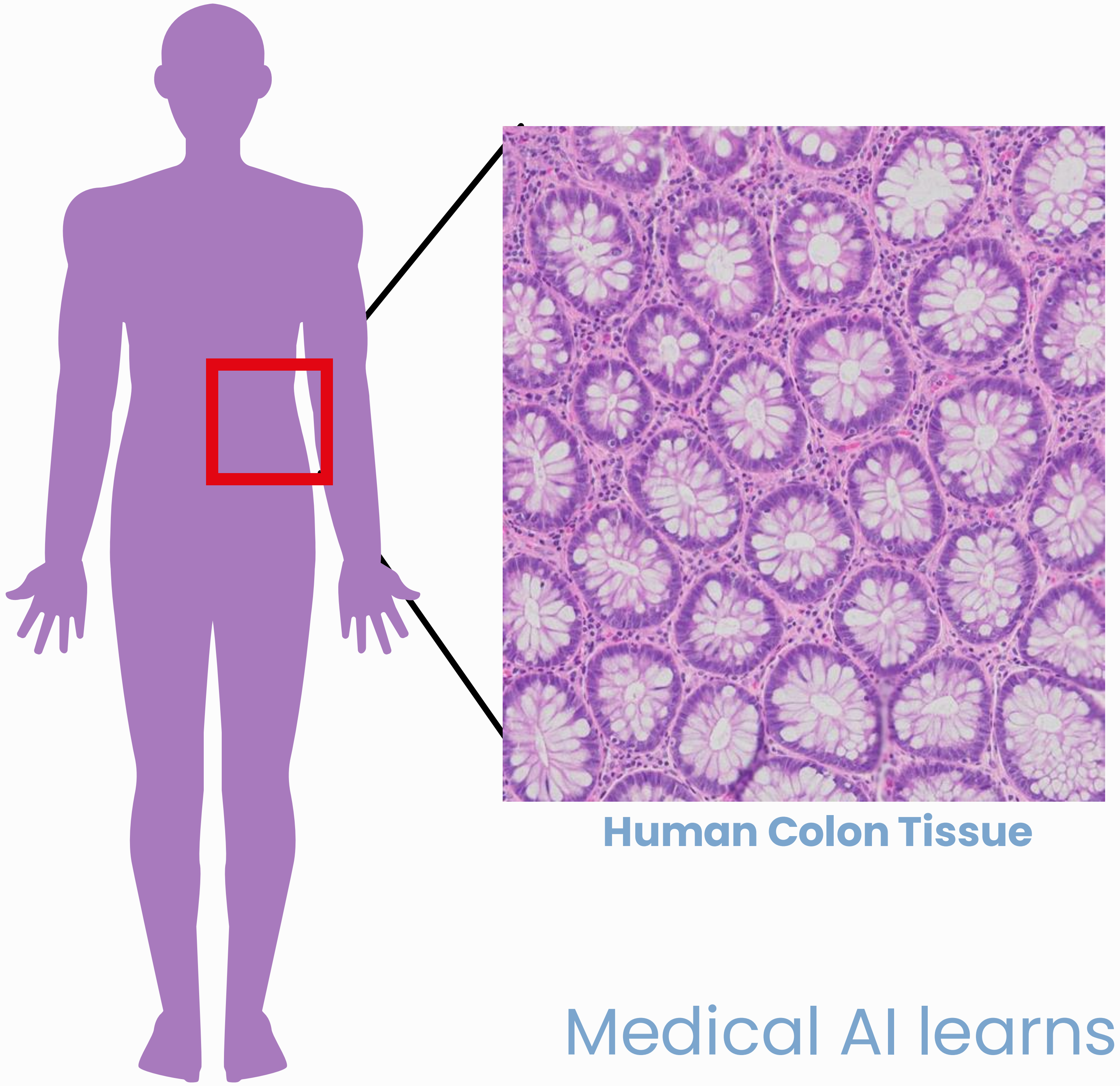
Time-intensive,
requires training,
creates bottleneck.



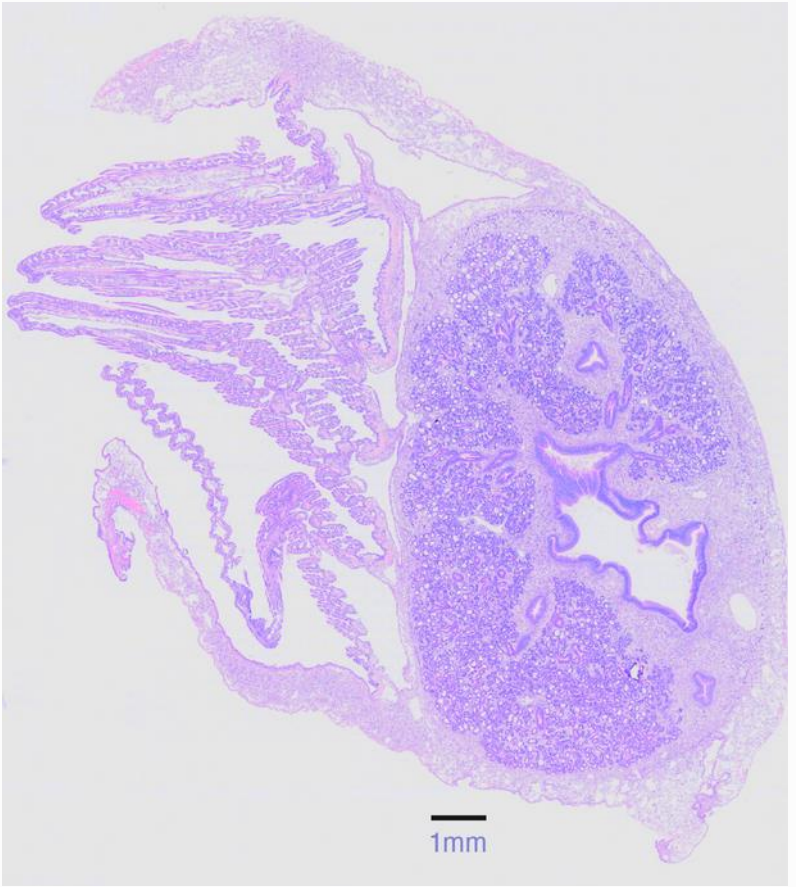
Manual Review by Pathologist

A Fundamental Domain Shift

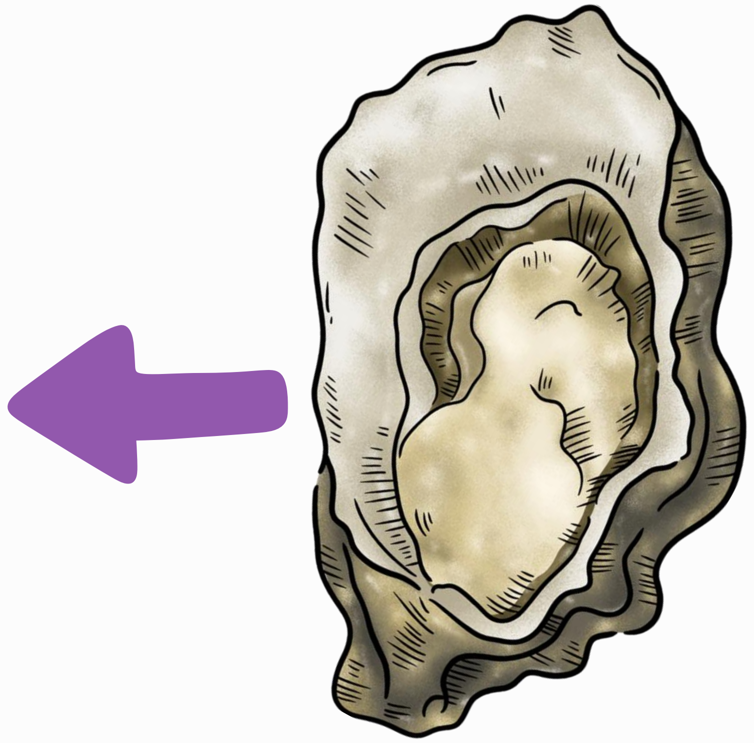
Why Medical AI Models Fail



Human Colon Tissue



Oyster Tissue



Medical AI learns from **small tissue biopsies**.
Our project analyzes a **whole organism** cross-section.

The Data Challenge

Building a High-Quality Dataset from Scratch

The Paradox: To build a tool that saves expert time, we first need a large investment of that same limited expert time.

The Strategy

A Human-in-the-Loop Workflow for Pathology

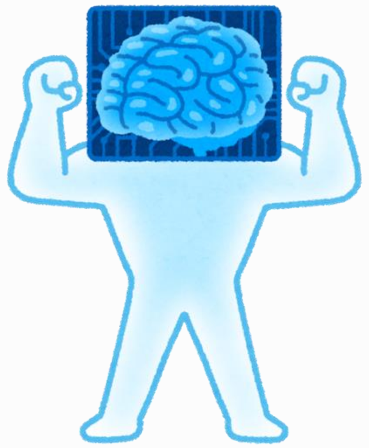
Stage 1: AI Pre-Analysis

The pipeline performs a first-pass analysis, generating initial annotations on a new slide.

Stage 2: Expert Review

A pathologist quickly validates and corrects the AI's suggestions, saving valuable time.

Stage 3: Model Retraining



The validated annotations are used to improve the AI model, making it smarter with each cycle.

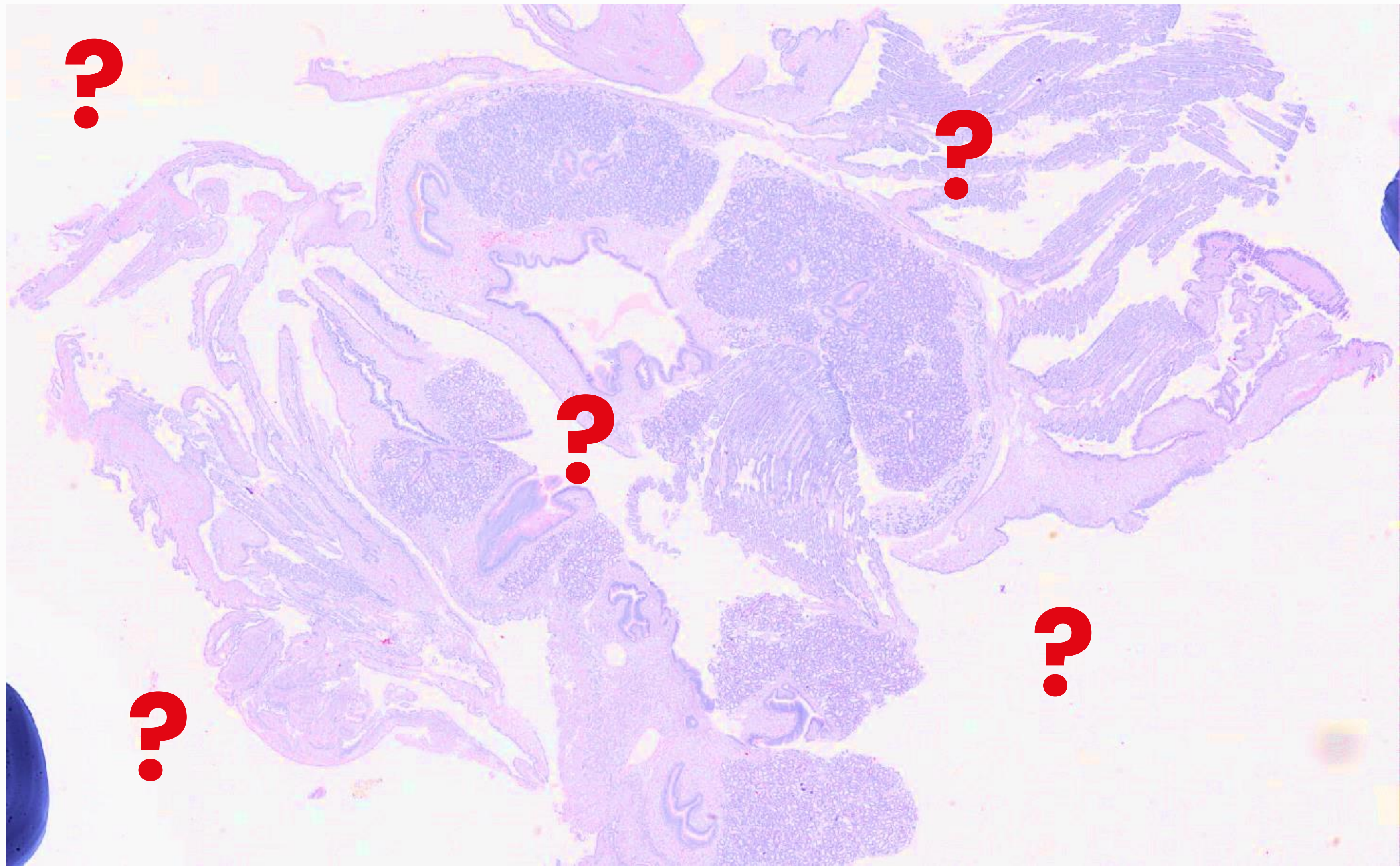
The First Hurdle

Robustly Separating Oyster Sections

Challenge: How can we create a single, automated process that works reliably for both of these cases without manual intervention?

The Ideal Case

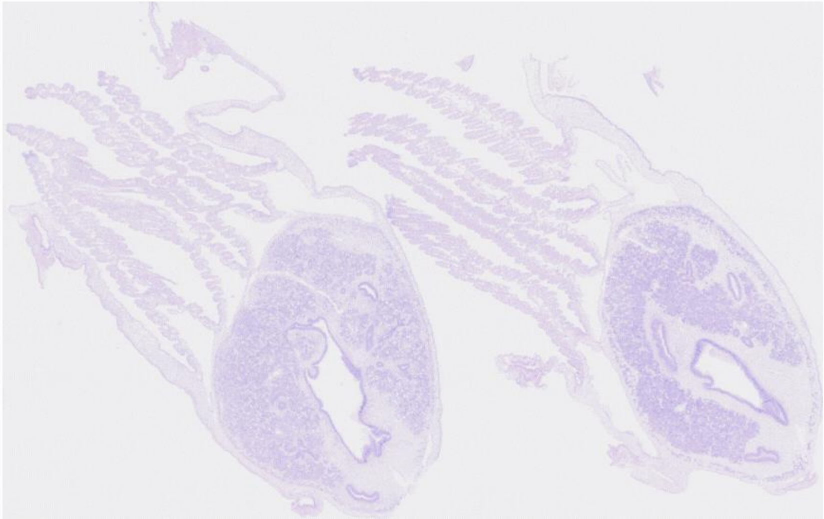
The Common Reality



Systematic Investigation

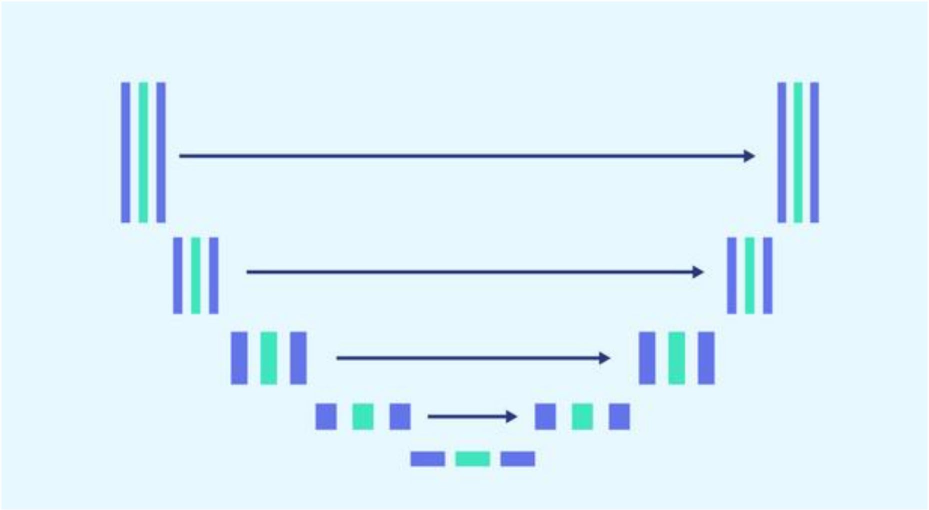
Goal: To find a robust, automated method that best replicates expert annotations.

Classical CV



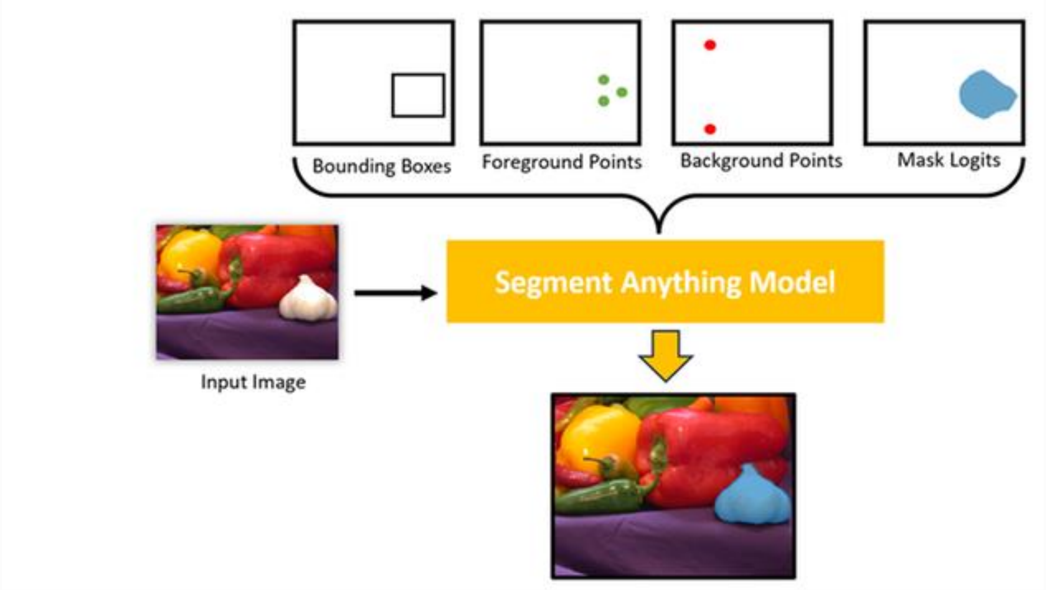
Method: Adaptive Thresholding & Watershed Algorithm.
Goal: Find a rule-based solution.

Supervised ML



Method: Train a U-Net model on 30 annotated slides.
Goal: Test a standard, data-driven approach.

Foundation Model

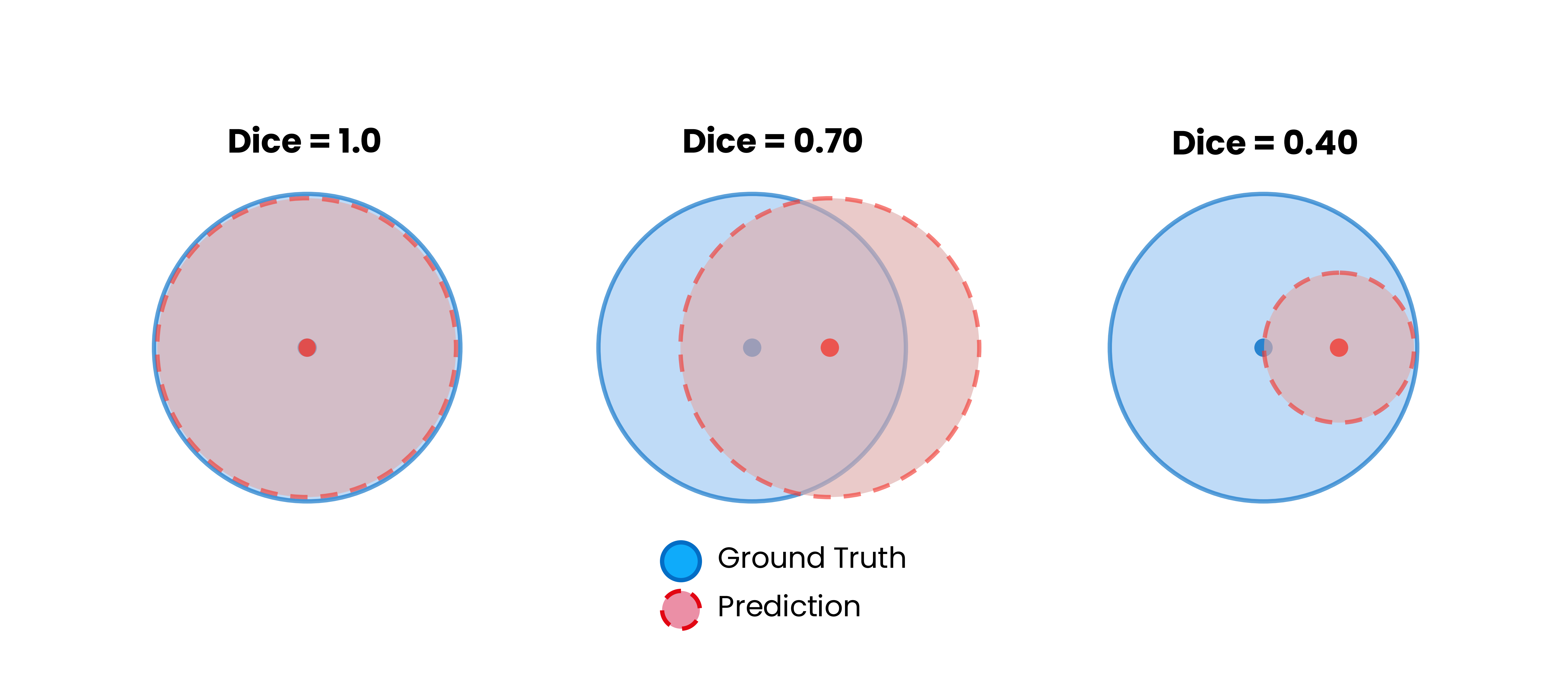


Method: Use a massive, pre-trained general model (SAM) guided by simple prompts.
Goal: Leverage a SOTA general model to overcome data limitations.

Key Findings & The Winning Approach

Approach	Methodology	Key Finding	Avg. Dice Score
Classical CV	Thresholding & Watershed	**Brittle & Parameter-Sensitive.** Fails on complex slides and requires constant manual tuning.	0.7463 ± 0.1287
U-Net	Supervised Deep Learning	**Data-Starved.** The dataset is too small (60 oysters), leading to poor generalization.	0.3801 ± 0.1104
SAM	Promptable Foundation Model	**Most Robust & Data-Efficient.** Intelligently handles fragmentation and leverages pre-trained knowledge.	0.8846 ±0.0600

Key Findings & The Winning Approach



SAM	**Promptable Foundation Model**	**Most Robust & Data-Efficient. Intelligently handles fragmentation and leverages pre-trained knowledge.**	**0.8846 ±0.0600**

Key Findings & The Winning Approach

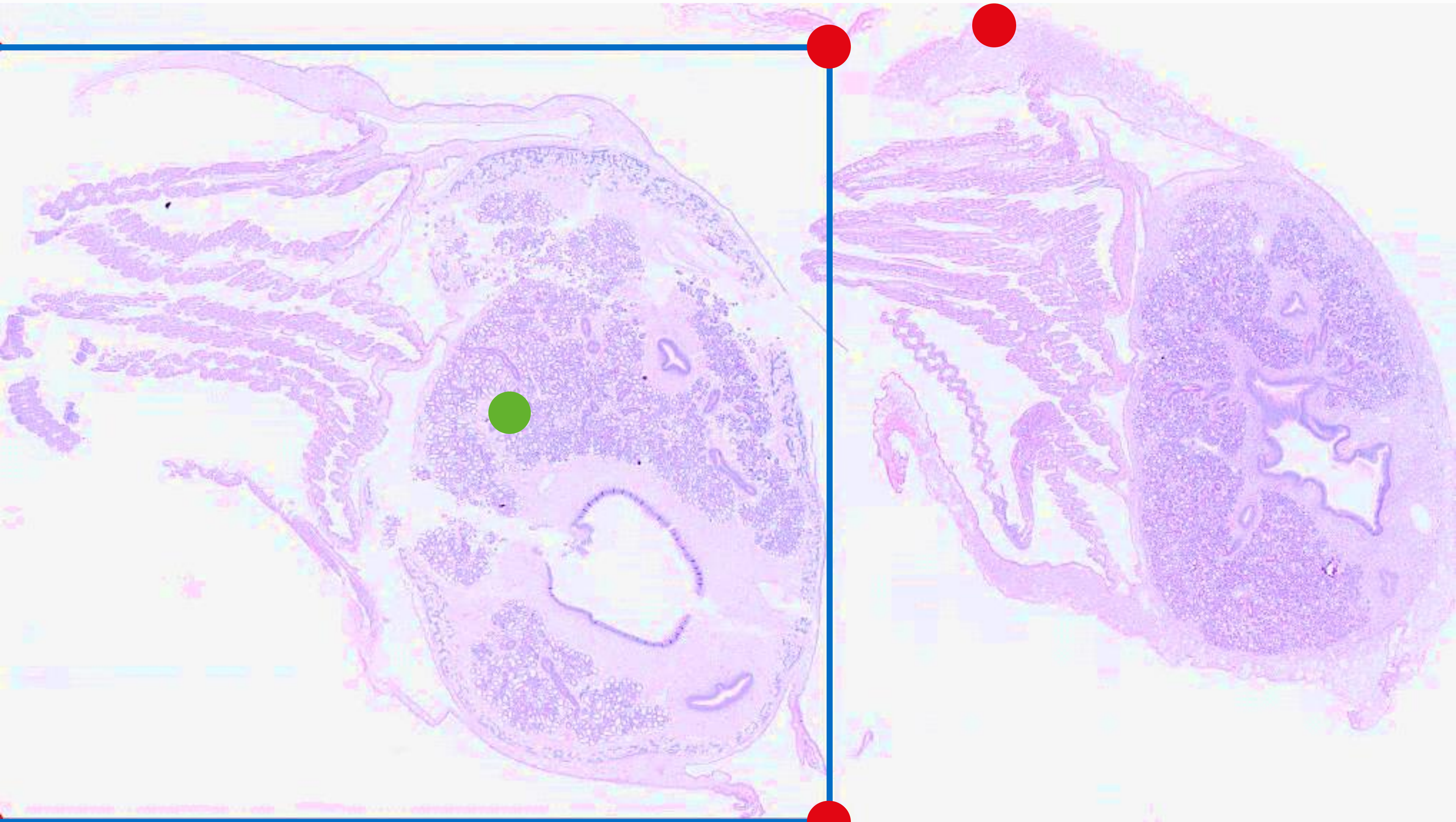
Approach	Methodology	Key Finding	Avg. Dice Score
Classical CV	Thresholding & Watershed	**Brittle & Parameter-Sensitive.** Fails on complex slides and requires constant manual tuning.	0.7463 ± 0.1287
U-Net	Supervised Deep Learning	**Data Starved.** The model was limited by the small dataset, leading to poor generalization.	0.3801 ± 0.1104
SAM	Promptable Foundation Model	**Most Robust & Data-Efficient.** Intelligently handles fragmentation and leverages pre-trained knowledge.	0.8846 ±0.0600

The Solution

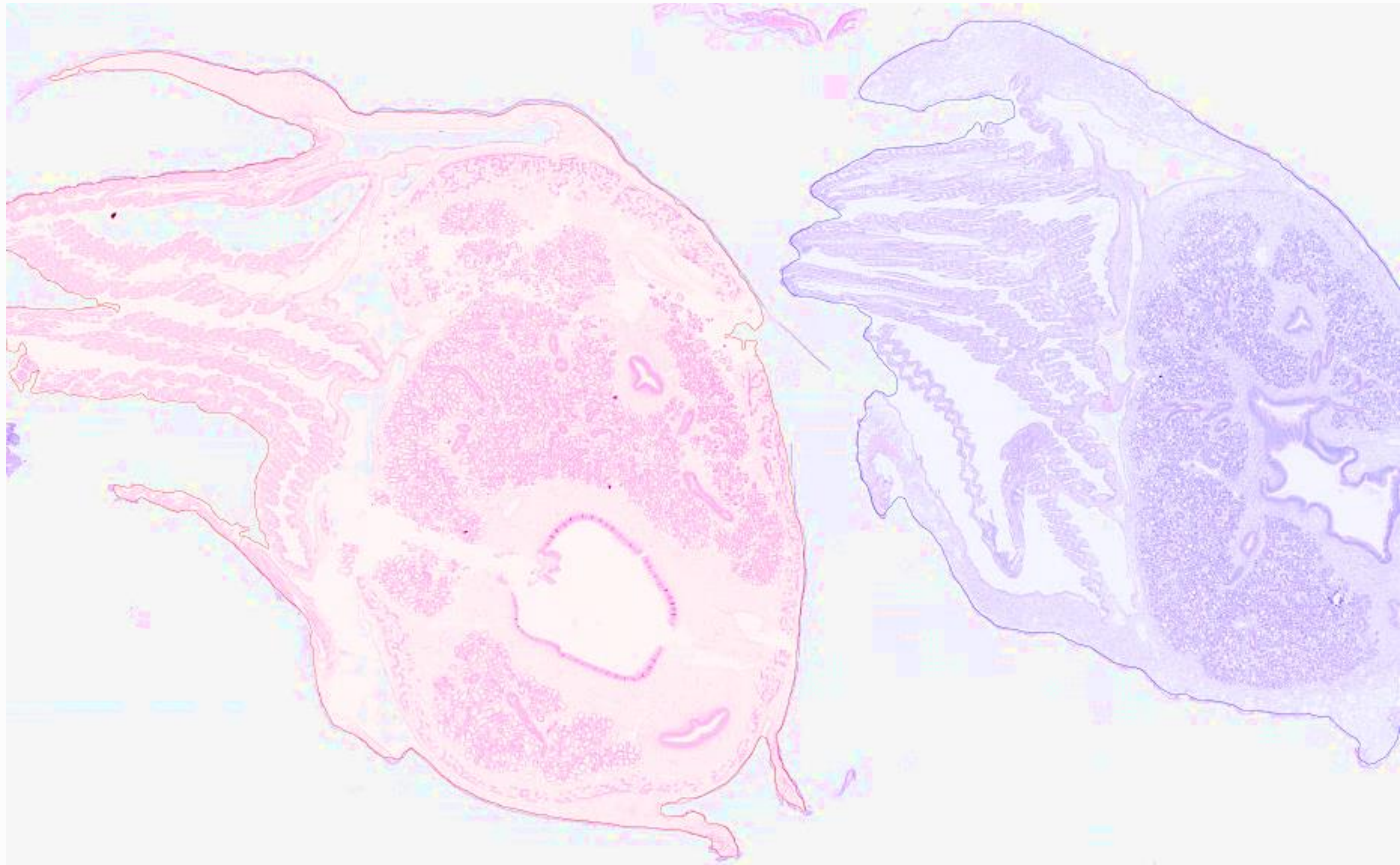
A Hybrid Pipeline for Intelligent Prompting

This hybrid strategy uses a simple, reliable method to answer 'where,' and the powerful foundation model to answer 'what.'

Step 1: Generate Robust Prompts (CV)



Step 2: Precise Segmentation (SAM)



The Solution

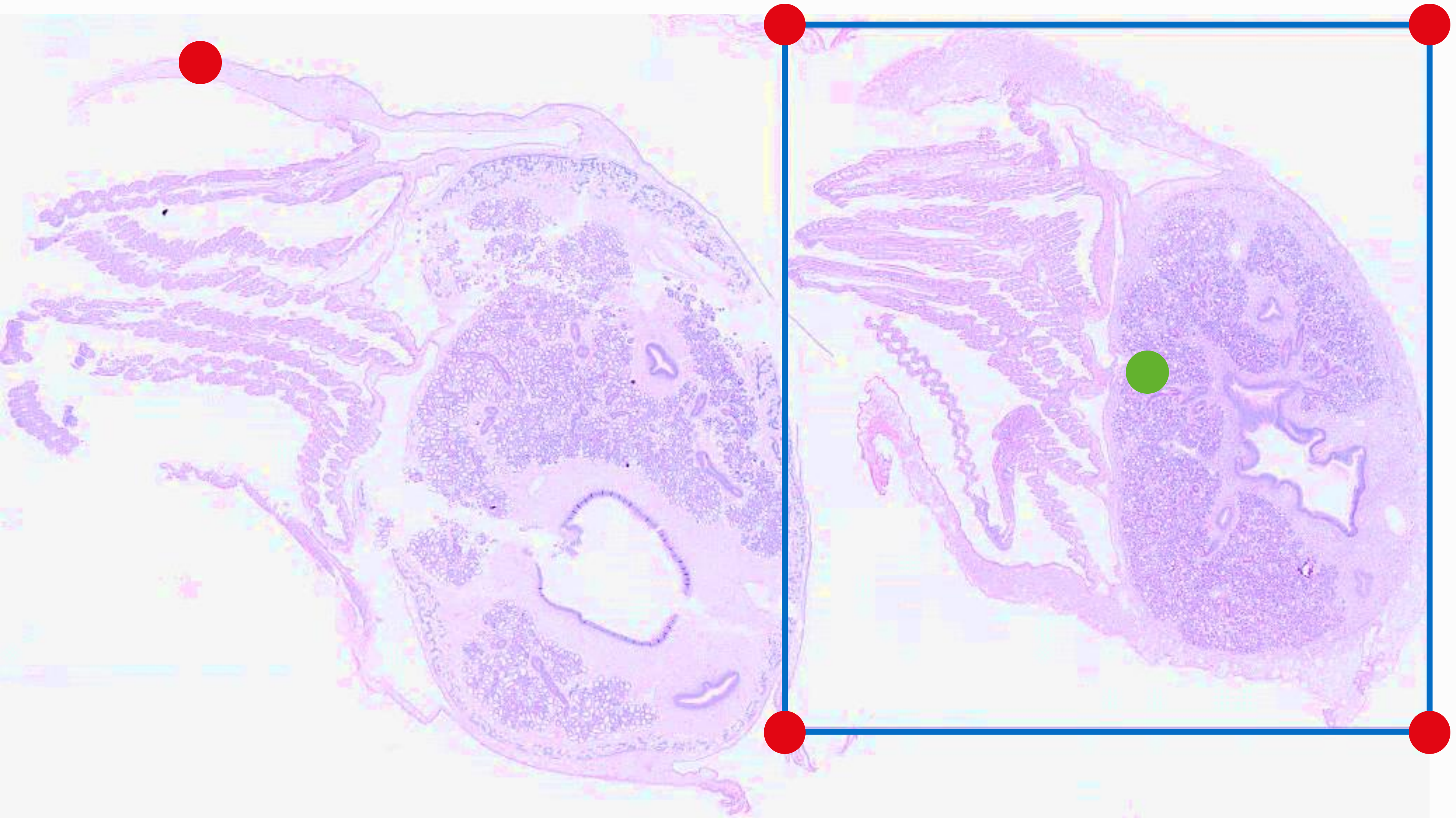
A Hybrid Pipeline for Intelligent Prompting

This hybrid strategy uses a simple, reliable method to answer 'where,' and the powerful foundation model to answer 'what.'

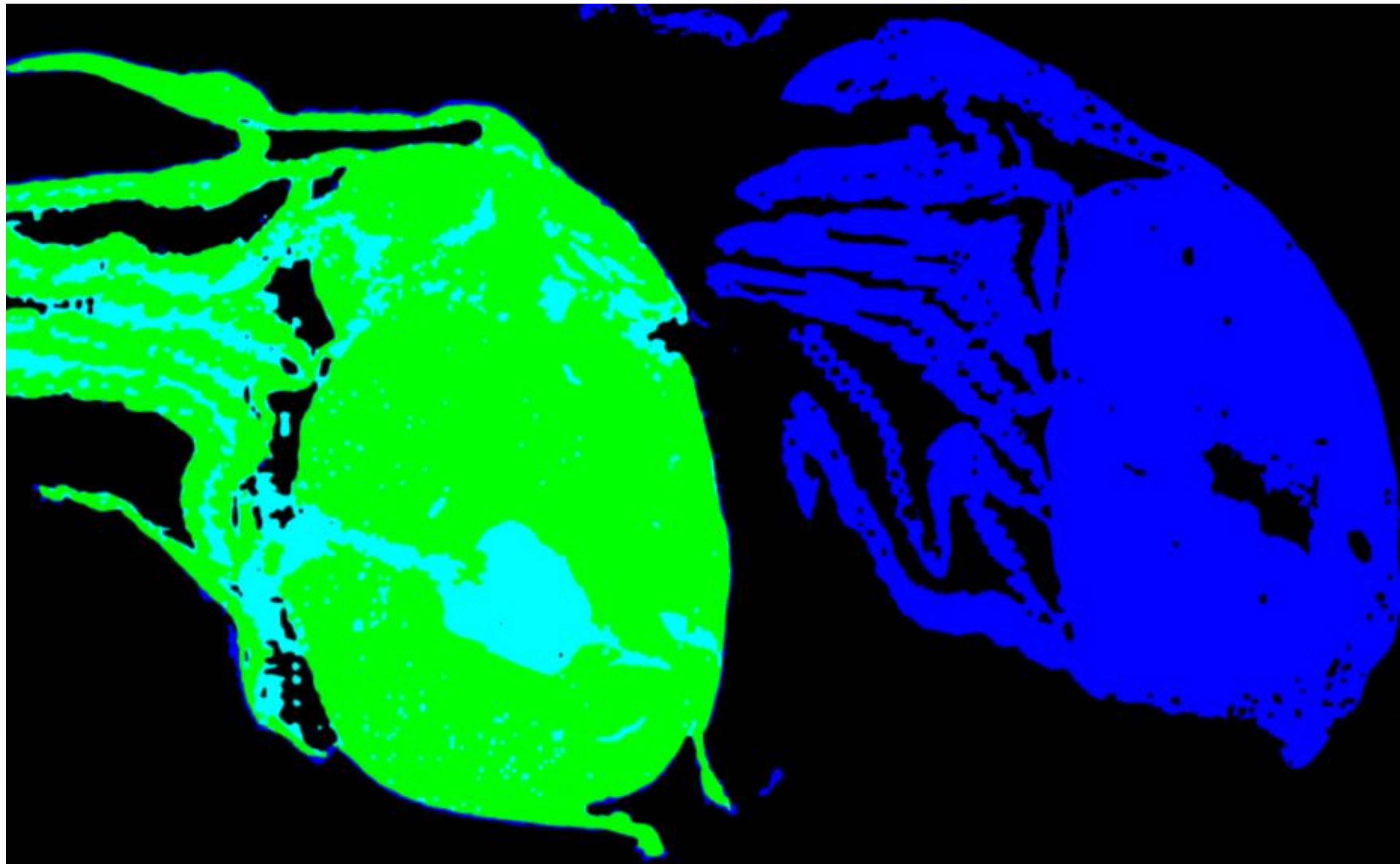
SAM's Guess

Expert Annotation

Step 1: Generate Robust Prompts (CV)



Step 2: Precise Segmentation (SAM)

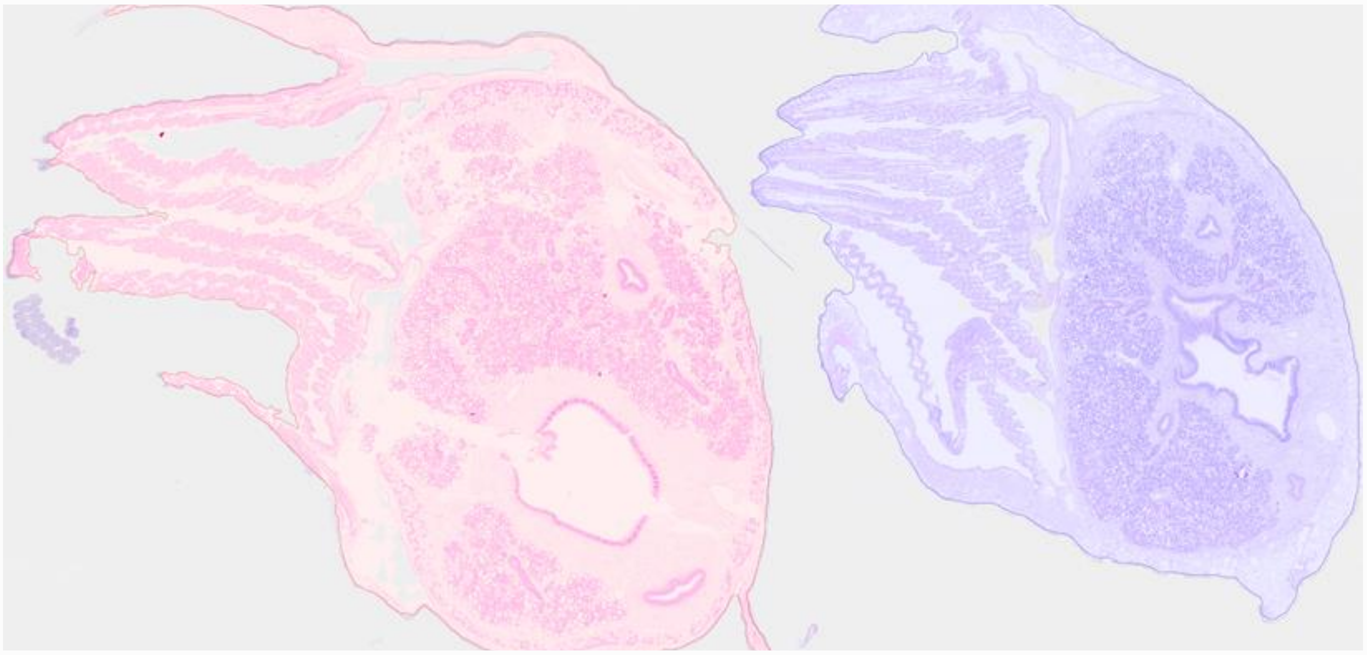


Performance Analysis:
Quantitative & Qualitative Results

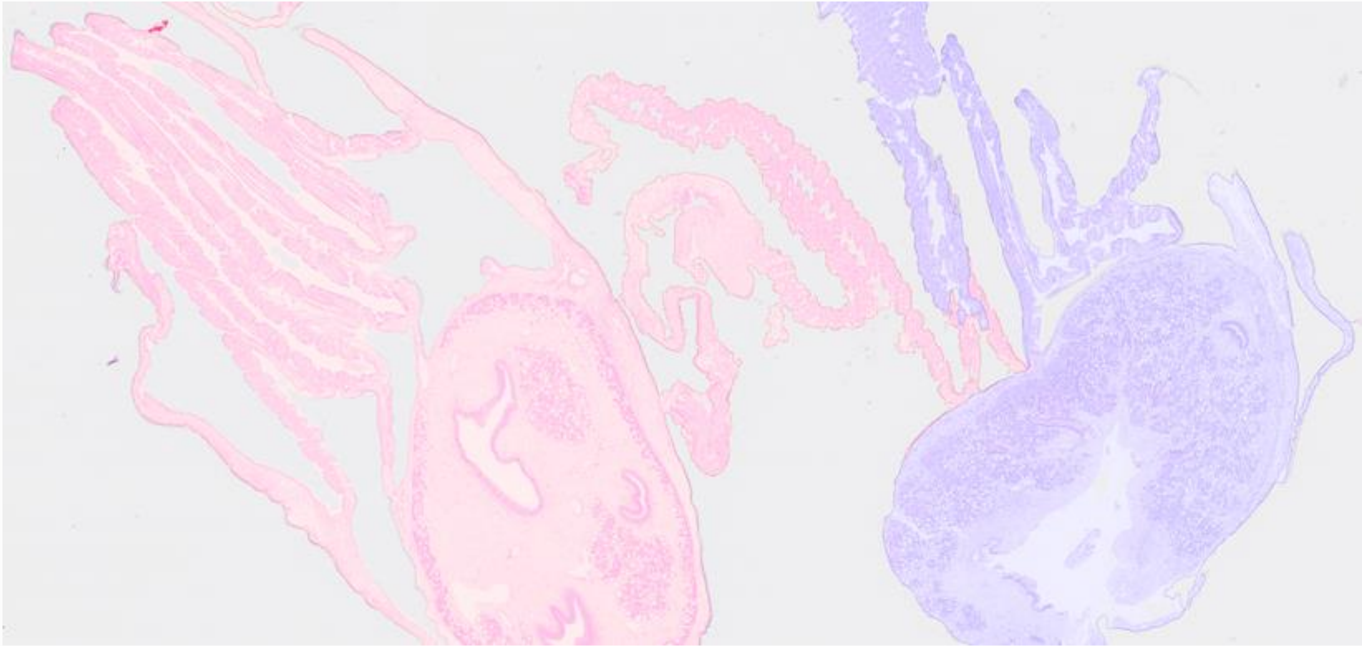
Quantitative Results (Avg. Across 30 Slides)

Metric	Score	Interpretation
Dice Score	0.8846 ±0.0600	Strong overlap; good proxy for boundary accuracy.
IoU (Jaccard)	0.8211 ± 0.0776	Good area-based overlap.
J&F Score	0.8528 ± 0.0685	Strong overall score, comparable to benchmarks.

Qualitative Results



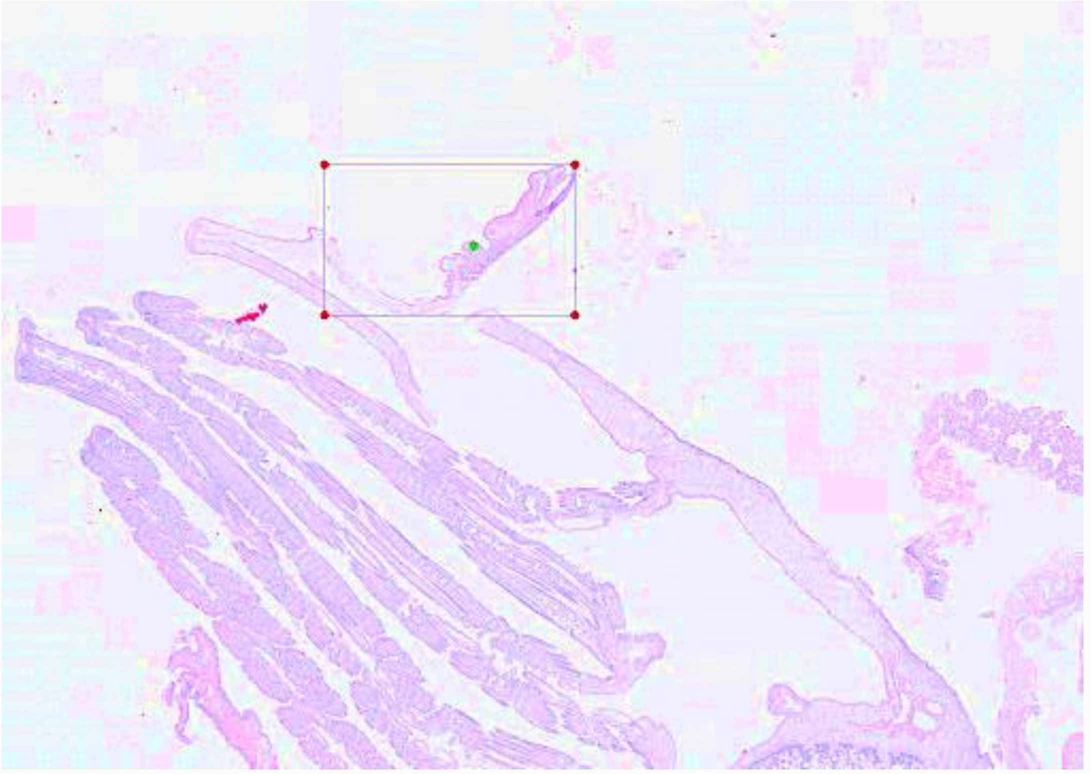
Success Case: Clean separation and tight boundaries on a well-defined slide.



Challenging Case: The model struggles when oysters are intertwined, showing the limits of the current prompter.

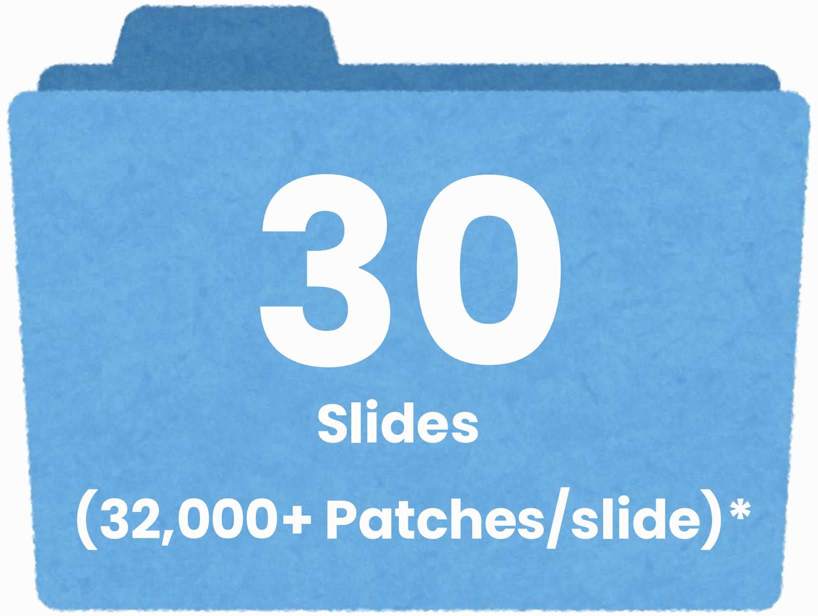
Current Limitations & Diagnosis

Brittle Prompt Generation



The current CV-based prompter is the weakest link and can fail on complex slides.

Small Dataset Size



While 30 slides generate hundreds of thousands of training patches, they don't capture the full range of biological variability needed for a production-ready model.

These limitations are not roadblocks; they are a clear and actionable guide for what to do next.

The Path Forward: A Phased Plan

Medium-Term

Large-Scale Data Annotation (Next 1-2 Months)

Objective: Expand the dataset to capture the full range of biological and technical variability.

Action: Utilize the complete HIL pipeline. Use our robust SAM segmenter to pre-annotate 50-100+ new slides for rapid expert validation.

Long-Term

Train the Production MSX Detector

Objective: Fulfill the project's primary mission of accelerating MSX diagnosis.

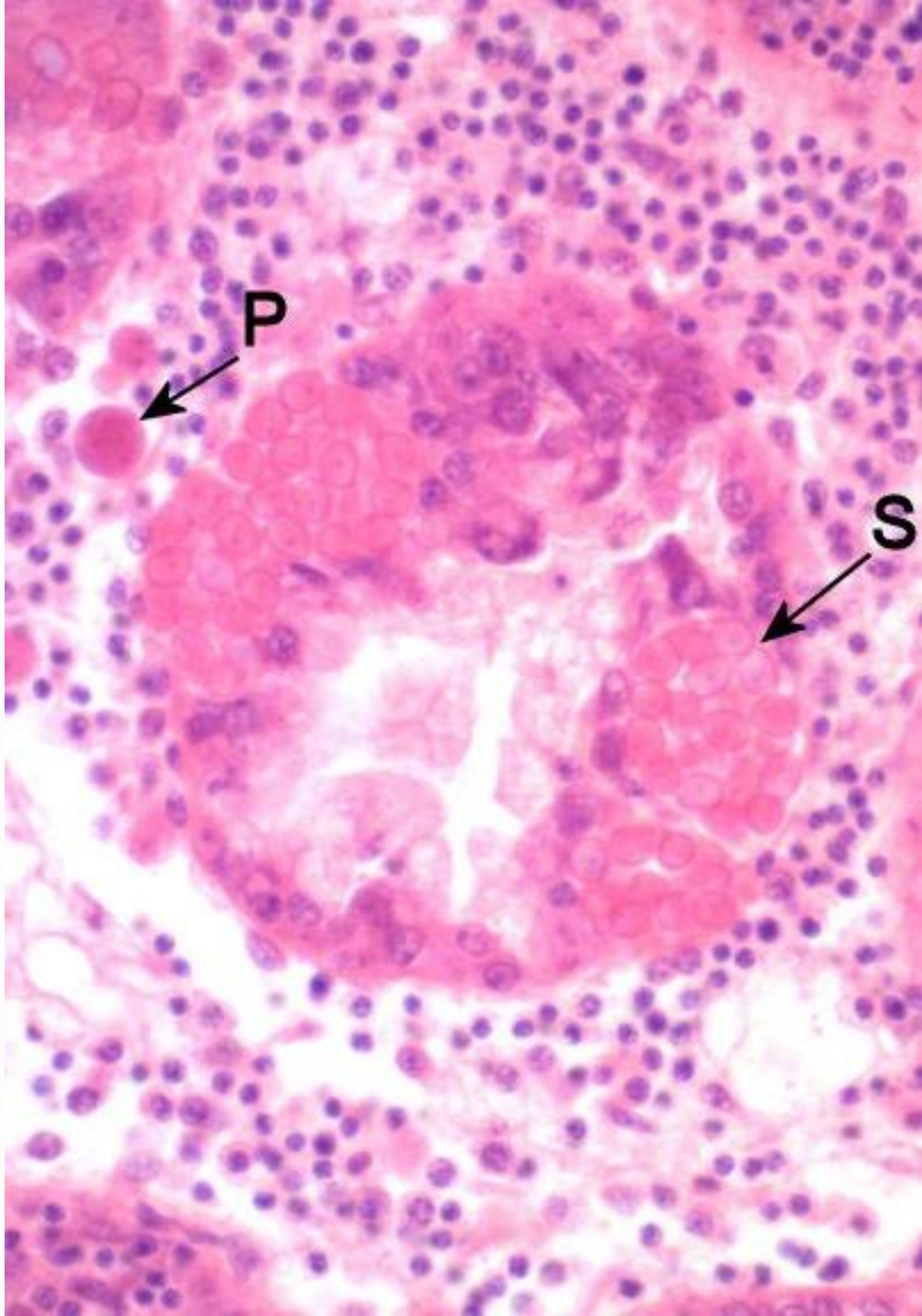
Action: With the segmentation problem solved and a large dataset in place, proceed with training a high-performance YOLO-based MSX detector.

Conclusion & Questions

Built a Complete HIL Pipeline
Successfully engineered and validated a complete, end-to-end Human-in-the-Loop pipeline for processing oyster histology slides.

Solved the Segmentation Challenge
Systematically investigated multiple methods and proved that a hybrid SAM-based approach is the most effective solution for the critical oyster segmentation task.

Established a Clear Path to Success
Have established a strong baseline performance, diagnosed the current limitations, and defined a clear, data-driven, three-phase plan for developing a production-ready MSX detection model.



AUGUST 2025

Thank You

BY JONATHAN E. ZUL LUNA