

AI-Assisted Annotations for Histology

Contents

Executive summary	2
Key project outcomes	2
Key learning for the team around using AI for developing teaching and learning	2
Project introduction	3
Background and context	3
Objectives	3
Scope	4
Tools and technologies	4
Dataset	4
Methodologies	4
Collaboration	5
Project outcomes and findings	6
Evaluation results	6
Lessons learned	6
Challenges	6
Key takeaways	7
How the Trained AI Model Will Be Used in CSlide V2	7
Advice for teams	8
Appendices	9
Evaluation results	9
Example 1: Good performance of the Model	9
Example 2: Good performance	11
Example 3: Excellent performance	11
Example 4: Moderate performance	12
Example 5: Poor performance	13
Example 5: Poor performance	14

Executive summary

Key project outcomes

Our project successfully demonstrated the potential of AI tools in the exploration and annotation of digital histology slides. This initiative has laid the groundwork for the development of CSlide V2, which will feature enhancements powered by our tested AI models. The project aimed to detect cells in histological slides, outline their boundaries, and classify them. We found that combining different AI techniques was generally effective for identifying and grouping cells, though classification of various structures proved less reliable. The three primary techniques employed were clustering, segmentation, and classification, each rigorously tested for performance and practical use. Our evaluation focused on qualitative observations. The segmentation pipeline accurately identified cellular structures by estimating pixel-level probabilities, which we used to generate polygonal outlines of cells and other features. Clustering and classification algorithms showed a reasonable ability to locate and group these. Looking ahead, we hope to expand this clustering approach to support efficient categorisation of cell types and tissue structures. This work offers a template for future improvements to histology teaching and workflows for both students and staff.

Key learning for the team around using AI for developing teaching and learning

Collaborating across technical and educational disciplines has provided a valuable learning opportunity for the entire team. Academic staff contributed essential domain knowledge that guided important project decisions, particularly in defining key annotation requirements. Their hands-on involvement, including manually annotating regions of interest on digital slides, was vital in supporting the efforts of research software engineers and medical science learning technologists (MSDLT). This interdisciplinary collaboration enabled the development and evaluation of algorithms, with team members jointly reviewing outputs to ensure alignment with the project's goals.

The project outcomes have shed light on both the potential and limitations of using AI to annotate histological slides for educational purposes. Several important factors came to light, such as the impact of tissue sample size and the need for accurately defined polygonal outlines around cellular structures. Although the segmentation and clustering models demonstrated encouraging performance, issues remain—such as missed cells and incorrect identification of non-cell components (eg nuclei or clustered structures) as cells.

One of the major challenges was the variability in AI model performance across different tissue types. While some tissues produced strong results, others were more problematic. This inconsistency may stem from differences in sectioning angles, structural complexity, the presence of diverse cell types, and artefacts of sample preparation.

Looking forward, the project will focus on scaling up the approach and gaining deeper insights into how AI models behave across diverse tissues.

Additionally, efforts will be directed at enhancing the workflow by integrating interactive components—such as an enhanced annotating and quiz feature in future—within the CSlide platform to further support teaching and learning in histology.

Project introduction

Background and context

The study of cells and tissues is a fundamental element of the medicine (primarily years 1, 2 & 4) and biomedical science (primarily year 1) courses. Understanding the microscopic structure of tissues is vital for grasping cellular physiology and function. To facilitate this learning, students engage in histology practicals led by instructors using microscopes, and CSlide—a digital platform developed by MSDLT.

CSlide hosts a wide array of scanned histological slides, enabling remote access and supporting distance learning, self-study, and review of practical material. CSlide's interactive tools allow students to zoom and pan slides. However, the complexity of labelling slides in the current system means that this feature is not used.

To address this, an AI-assisted histology project was launched to evaluate how artificial intelligence could improve the annotation process by segmenting and outlining specific cells or clusters. The goal is to ease the creation of annotated resources, reducing workload for teaching staff and offering students a more autonomous and engaging learning experience.

Objectives

The main objective is to enhance histology education by using AI to support the annotation and exploration of digital tissue slides. This includes automating identification, segmentation, clustering, and cell labelling. The project has assessed multiple AI models to ensure accuracy in segmentation and clustering, aiming to develop a flexible tool suitable for various tissue types.

Scope

This project focussed on building and deploying an AI-driven pipeline for automating key processes: image tiling, filtering, segmentation, clustering, and annotation. Tools like Cellpose (for segmentation) and a combination of ResNet-50, UMAP, and DBScan (for clustering) were used to manage the very large datasets efficiently. Outputs were integrated into the developing CSlide v2 platform. The project did not include developing new classification models, due to limitations of large language models, or real-time lab-based annotations, because of the time taken for processing and because full automation without human validation was out of current scope.

Tools and technologies

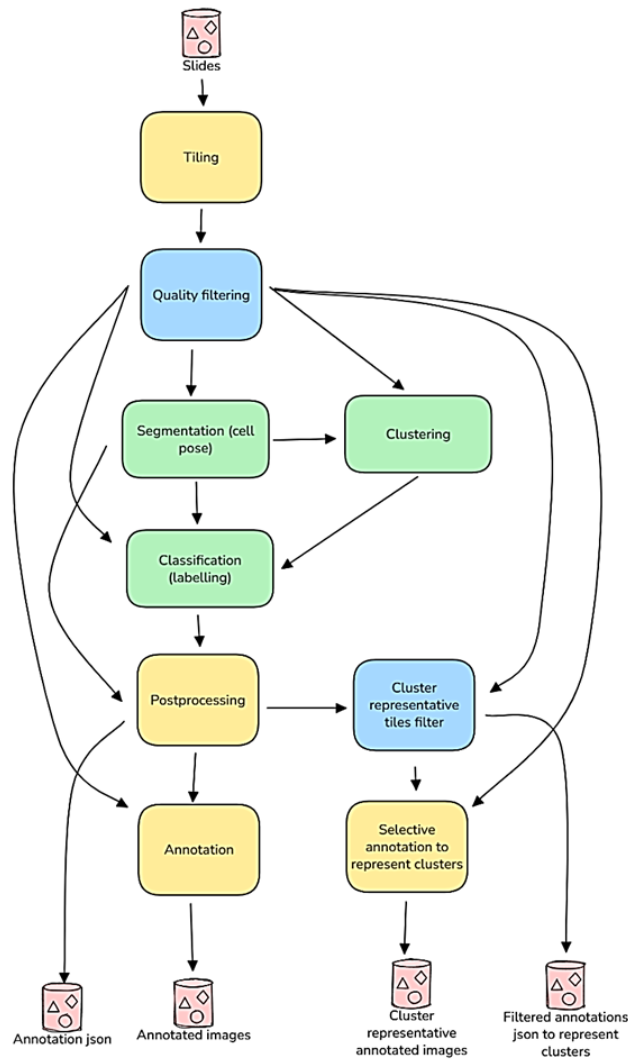
Dataset

- Digital histology slides from the CSlide platform.
- A limited number of manually annotated slides served as a benchmark for evaluating AI performance qualitatively.

Methodologies

The AI pipeline starts with tiling high-resolution slide images into 512×512 pixel sections. Tiles are filtered to exclude low-quality images. Cellpose performs segmentation, generating cell probability maps. Segmented results from all tiles are grouped using clustering algorithms, distinguishing similar cell types (note: AI 'clustering' differs from anatomical clustered-cell structures). Classification (ie naming cells) was performed on representative samples, but poor performance meant that the emphasis remained on segmentation and clustering. A JSON output consolidates all results – this is then fed to CSlide v2 for viewing.

Below is the simplified workflow demonstrating the step-by-step process involved in the annotation.



Collaboration

This interdisciplinary effort involved academic experts, a research software engineer (RSE), and MSDLT.

Academics provided critical histological input and manually annotated slides, helping evaluate AI outputs and guide model expectations.

The RSE developed and tuned the AI pipeline—managing tiling, segmentation, clustering, and evaluation—adapting it to diverse tissue complexities despite limited ground truth data.

MSDLT connected AI outputs to pedagogy, recommending tools like OpenSeadragon for slide interaction, and segmentation models such as SAM, MedSAM, and MedSAM-2 for advanced annotation support. They integrated outputs into CSlide v2 to ensure for viewing.

A continuous feedback loop allowed for refinement across teams. Academics reviewed AI results, engineers improved model parameters, and technologists ensured learning usability.

This project highlights the benefits of interdisciplinary collaboration, balancing education, technical development, and pedagogical implementation—laying the groundwork for scalable, AI-enhanced histology resources.

Project outcomes and findings

Evaluation results

The evaluation of AI-assisted cell segmentation was conducted across a diverse array of histological tissues, with different cellular architectures including skeletal muscle, pancreas, adrenal gland, thyroid gland, and posterior pituitary. Overall, the results demonstrate promising segmentation capabilities, especially in tissues with moderate to low cell density and well-defined nuclear morphology, while also exposing challenges in more complex or densely cellular regions. For example, the pancreas showed high segmentation accuracy in exocrine regions; the adrenal gland performed well in the zona glomerulosa and fasciculata; the skeletal muscle tissue demonstrated good segmentation of a muscle cell; the thyroid gland showed partial success with follicular cells but struggled with flattened epithelium and artefacts; and the posterior pituitary exhibited poor performance, with the model outlining empty fibrous regions instead of cells. These results highlight that the segmentation model performs best in moderately cellular tissues with clear nuclear morphology and adequate spatial separation, but has limitations in densely packed, fibrous, or architecturally complex tissues. These findings underscore the importance of tissue-specific model tuning and the inclusion of diverse training samples to improve cross-tissue performance.

A detailed evaluation of the results is provided in the appendices below.

Lessons learned

Challenges

The diverse range of cell types and tissue structures meant that a single segmentation model or configuration worked well for some slides but underperformed on others. This suggests that adapting model selection or tuning configurations per slide may lead to improved segmentation accuracy.

Processing extremely large histology slides presented computational and time-related constraints. Since tissue often occupies only a portion of the full slide, strategies were implemented to manage this: slides were divided into 512x512 tiles, low-quality tiles (eg too bright, too dark, or lacking structure) were filtered out using brightness and edge metrics, and GPU-based processing was employed. Though tile sampling was rarely used, it remained an option.

Initial setup of the Azure environment for deploying and managing pipeline components posed significant technical hurdles.

Key takeaways

How the Trained AI Model Will Be Used in CSlide V2

The AI pipeline (comprising multiple models and processes) developed in this project will underpin the enhanced features of CSlide V2, offering benefits for both staff and students. CSlide V2 will be embedded in Canvas as an LTI tool.

For Staff: Streamlined Annotation

My MT25, CSlide V2 will allow easier labelling of slides which have been pre-processed by an AI pipeline:

- Tools for staff to:
 - Label all items in a cluster collectively
 - Edit, delete, or refine polygons as needed
 - Add new labels or regions manually

This reduces annotation time while retaining expert control and ensuring teaching resource quality.

The AI pipeline (currently fairly slow and requiring an RSE to run) needs further development to build on existing functionality:

- Detection and grouping of similar cells/structures
- Automated polygon generation around detected elements

In particular:

- testing whether we can identify, and then automatically or manually select models which perform well on the different types of tissue used in teaching
 - Whether we can integrate the AI pipeline into CSlide V2s associated with processing hundreds of slides
 - How we deal with the cost
 - Whether automated (albeit with human in the loop) labelling might become possible as models improve.

For Students: Interactive Learning Tools

By MT25, students will benefit from:

- Slides with AI-assisted labelled structures

Further work will be required to deliver:

- Embedded quizzes (eg “Label this structure” or “Click on a cell of type X”)

Early prototypes show labelled, zoomable slides with hover-to-reveal functionality, navigable by clicking cell types. Future enhancements may include arrows and interactive lists to guide users through slide content.

Self-assessment tools could include “Click on a cell of type X” or “Identify this highlighted cell,” using AI-generated polygons/staff labelling to support active learning.

Advice for teams

Teams exploring AI in histology should prioritize early collaboration among educators, technologists, and engineers. High-quality annotated datasets are essential. Models excel with skeletal muscle and pancreas but struggle with fibrous, artefactual, or sparsely populated tissues. A human-in-the-loop model supports refinement, and annotation tools should enable flexible, mixed-mode workflows. Embedding outputs into student platforms with interactive features enhances usability. Documenting and sharing tissue-specific insights is key to wider adoption.

Appendices

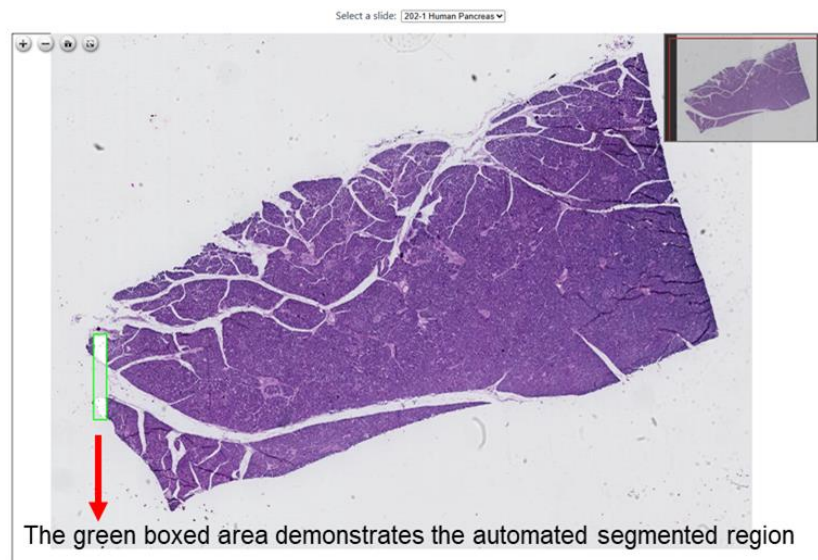
Evaluation results

Sample prototype accessible via: <https://learntech.medsci.ox.ac.uk/cslide-lite/>

Example 1: Good performance of the Model

Slide number: 202-1

Tissue: Human Pancreas



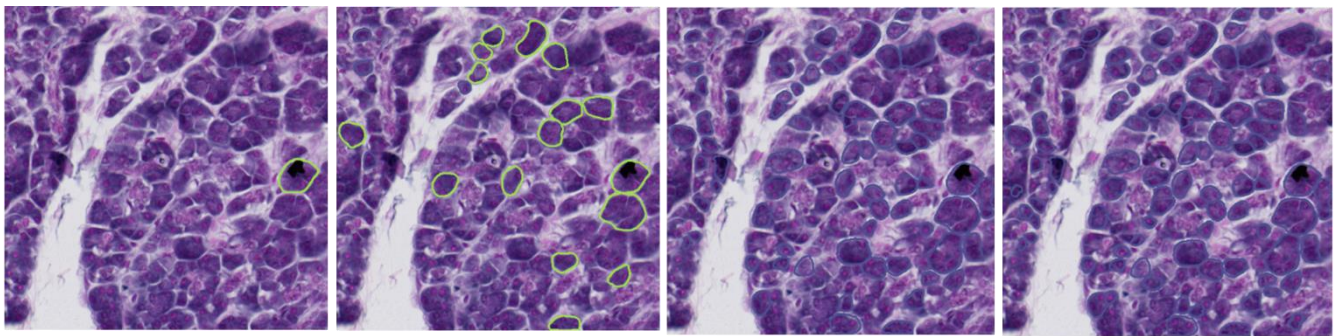
Two key tuning parameters were tested to refine the segmentation quality in our setting, probably threshold and flow threshold. Probability threshold determines the strictness for the model in terms of identifying a given pixel on an image as part of a cell, a higher value is expected to remove noise. Flow threshold determines the strictness of shape; a smaller value is expected to filter out noise.

Evaluation of the pancreatic tissue – AI model interpretation

Slide number: 202-1

Tissue: Human Pancreas (Exocrine component)

Segmentation Model: cellpose 3.1.1 (released Feb 2025)



Panel 1 prob=0, flow=.05

Panel 2 prob=1, flow=.05

Panel 3 prob=2, flow=.2

Panel 4 prob=2, flow=.4

'prob' refers the probability of a pixel on the image being part of a cell, 'flow' refers the shape adjuster - a lower value permits smoother cell shapes

Observation

Panel 1 No segmentation, the green outline shows the staining artefact.

Panel 2 Under segmentation, the model has detected mostly large, well-defined cells, missing smaller or tightly clustered acinar cells.

Panel 3 Intermediate segmentation (improved coverage)

Noticeably more cluster of cells are outlined, covering most of the tissue architecture.

The model begins to detect clusters of acinar cells more reliably, though some boundaries remain rough or imprecise.

Detection of smaller nuclei improves, indicating better sensitivity and possibly updated model parameters or training adjustments.

Panel 4 High fidelity segmentation

Shows fine, smooth segmentation outlines around nearly all visible cluster of cells.

Acinar cells, which are densely packed and sometimes irregular in shape, are clearly distinguished, demonstrating strong model generalization.

This output represents a robust segmentation result, suitable for further tasks like clustering, classification, or annotation for teaching and analysis.

Example 2: Good performance

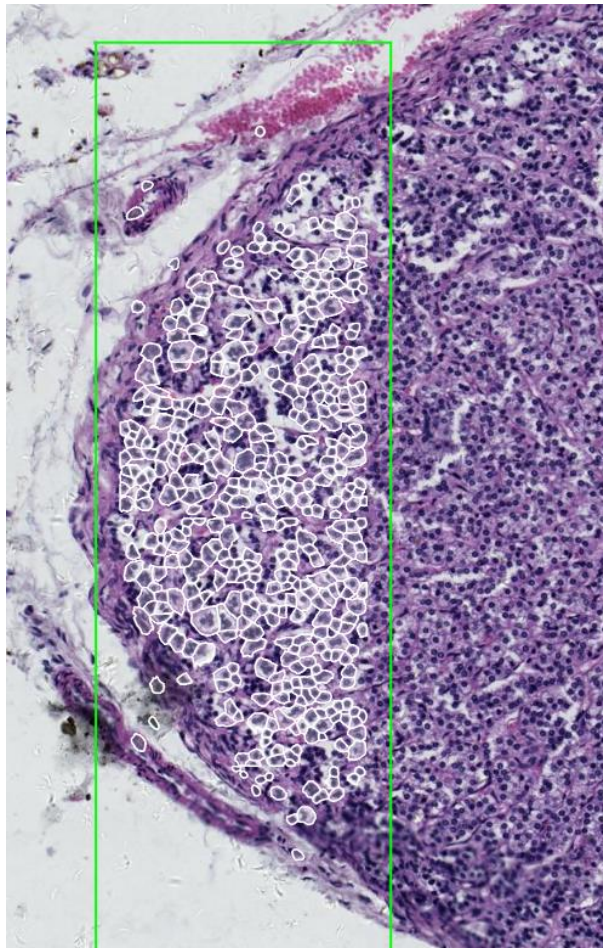
Slide number: 474

Tissue: Human Adrenal gland – Cortex part

Segmentation Model: cellpose 3.1.1 (released Feb 2025)

Observation

The adrenal gland is well organised layered structure, includes outer cortex and inner medulla, The segmentation model effectively identified cell boundaries in regions where cells are larger, more isolated, and well-spaced—typical of the glomerulosa and fasciculata. These segmented cells appear uniform in size and distribution, consistent with adrenal cortical cells, which tend to have clear borders and rounded nuclei. These areas exhibit enough contrast and regularity for the AI model to confidently apply polygon boundaries around individual cells.

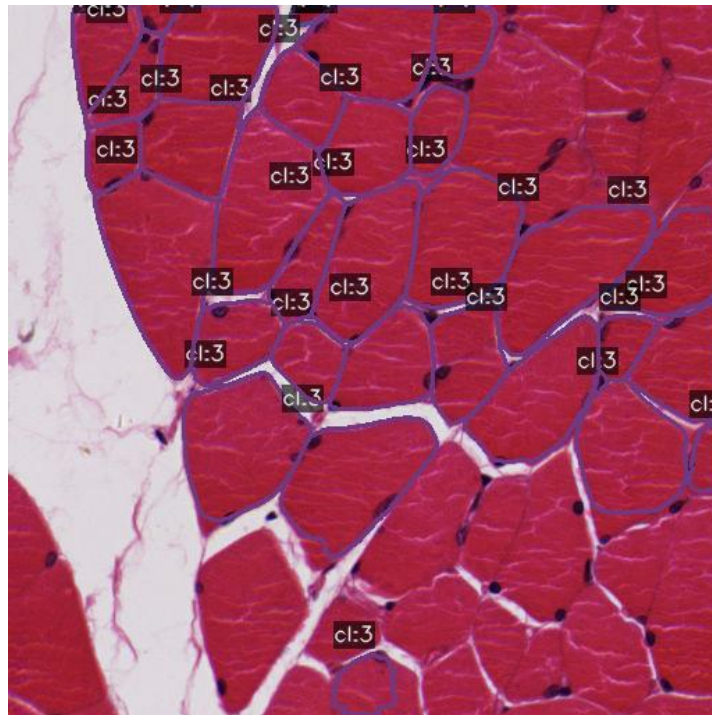


Example 3: Excellent performance

Slide number: 93

Tissue: Skeletal muscle

Segmentation Model: cellpose 4.0.5 (released 26 Jun 2025)



Observation: The skeletal muscle cells are multinucleated and peripherally located, this is unique feature of the tissue. The AI segmentation model successfully delineated individual muscle fibres and each segmented fibre is assigned a cluster ID [cl:3], indicating that the model has grouped these muscle fibres into a single morphological or cluster. Here the segmentation accuracy is high, with clear boundaries and consistent shapes reflecting the regular microstructure of the skeletal muscle in this transverse profile. The skeletal muscle as one of the strongest performing tissue types for this segmentation model.

Example 4: Moderate performance

Slidenummer:498

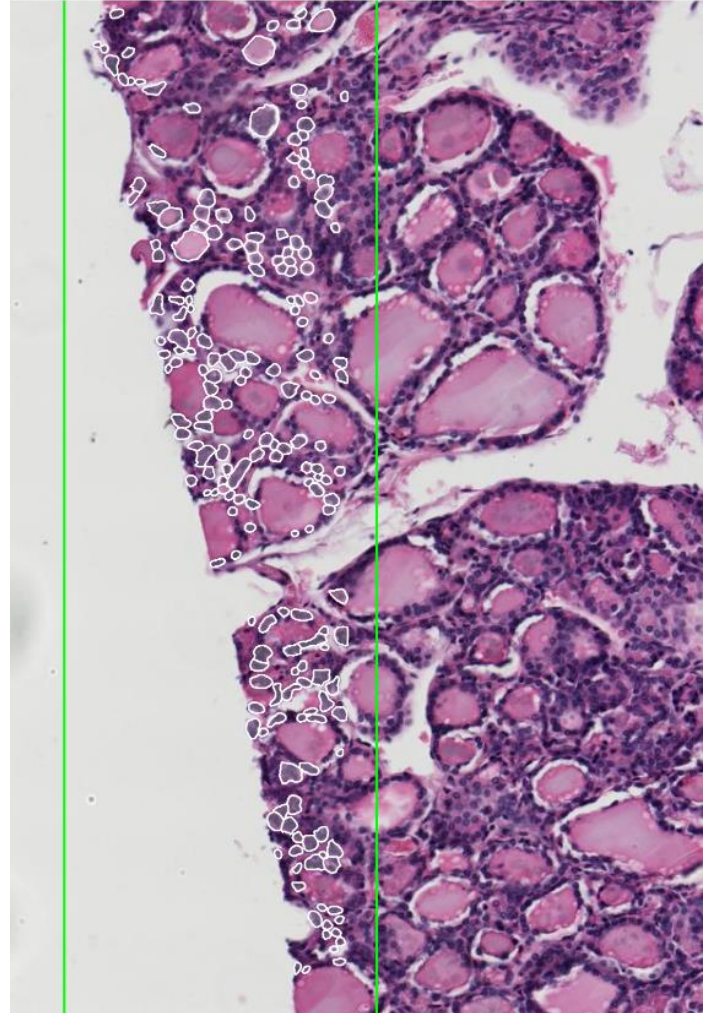
Tissue: Human Thyroid gland

Segmentation Model: cellpose 3.1.1 (released Feb 2025)

Observation:

The thyroid gland is composed of numerous spherical follicles lined by a single layer of follicular epithelial cells surrounding a central lumen filled with colloid. The surrounding stroma contains parafollicular (C) cells, blood vessels, and connective tissue.

The AI segmentation model applied to the thyroid gland image showed partial success, primarily detecting individual follicular epithelial nuclei along the edges of colloid-filled follicles in the left portion of the image. These cells, typically cuboidal or columnar, were outlined with polygon overlays, especially where nuclear boundaries were more distinct. However, the model missed many cells, particularly in regions with flattened or low cuboidal epithelium, where nuclear contrast is minimal.



Large acellular colloid areas dominated the tissue, reducing overall cell density and potentially affecting the model's focus during segmentation. Additionally, the model appeared to falsely segment some artefactual gaps near tissue edges. These limitations likely stem from a combination of factors: the subtle morphology of inactive follicular cells, the high variability in follicle size and shape, and the model's bias toward tissues with denser, more defined cellular structures. Overall, while the model performs reasonably well in favourable regions, it struggles to generalize across the full histoarchitectural spectrum of the thyroid, particularly in distinguishing cells surrounding variably sized colloid spaces.

Example 5: Poor performance

Slide number: 450

Tissue: OX Pituitary gland-
Neuroendocrine tissue

Model: Segmentation Model: cellpose
3.1.1 (released Feb 2025)

Observation: Polygon outlines generated around areas that appear empty or sparsely structured; no obvious cellular structures detected by the model in a region known to be rich in nerve fibers and supporting (glial) cells –posterior pituitary gland. The AI model likely failed to detect cells in the posterior pituitary region due to several key factors. First, the dominant presence of unmyelinated nerve fibers and sparse, poorly defined glial cells (pituicytes) provides limited structural contrast for the model, which is typically trained on tissues with well-defined nuclei and cell boundaries. The fibrous texture may have confused the model, leading it to outline empty inter-fiber spaces instead. Additionally, the staining used may emphasize connective or neural elements over cellular detail, reducing the visibility of actual cells. Finally, a lack of training data specific to neuroendocrine tissues like the posterior pituitary could have limited the model's ability to generalize effectively in this context.

