A pathologist–AI collaboration framework for enhancing diagnostic accuracies and efficiencies

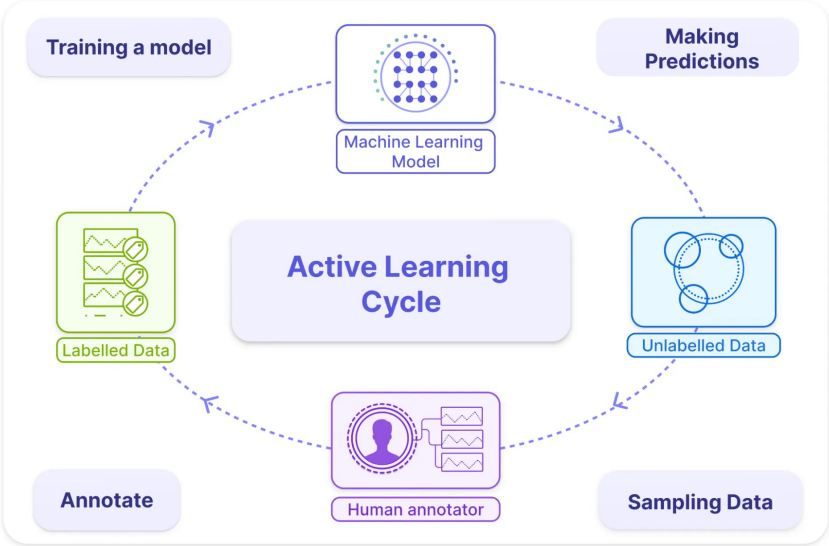
Zhi Huang, Eric Yang, Jeanne Shen, Dita Gratzinger , Frederick Eyerer, Kathleen S. Montine, Thomas J. Montine & James Zou

Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA.
Department of Biomedical Data Science, Stanford University School of Medicine, Stanford, CA, USA.

Active learning

- A special case of machine learning in which a learning algorithm can interactively query a human user (or some other information source), to label new data points with the desired outputs.

- The human user must possess knowledge/expertise in the problem domain



ENCORD

https://encord.com/blog/active-learning-machine-learning-guide/

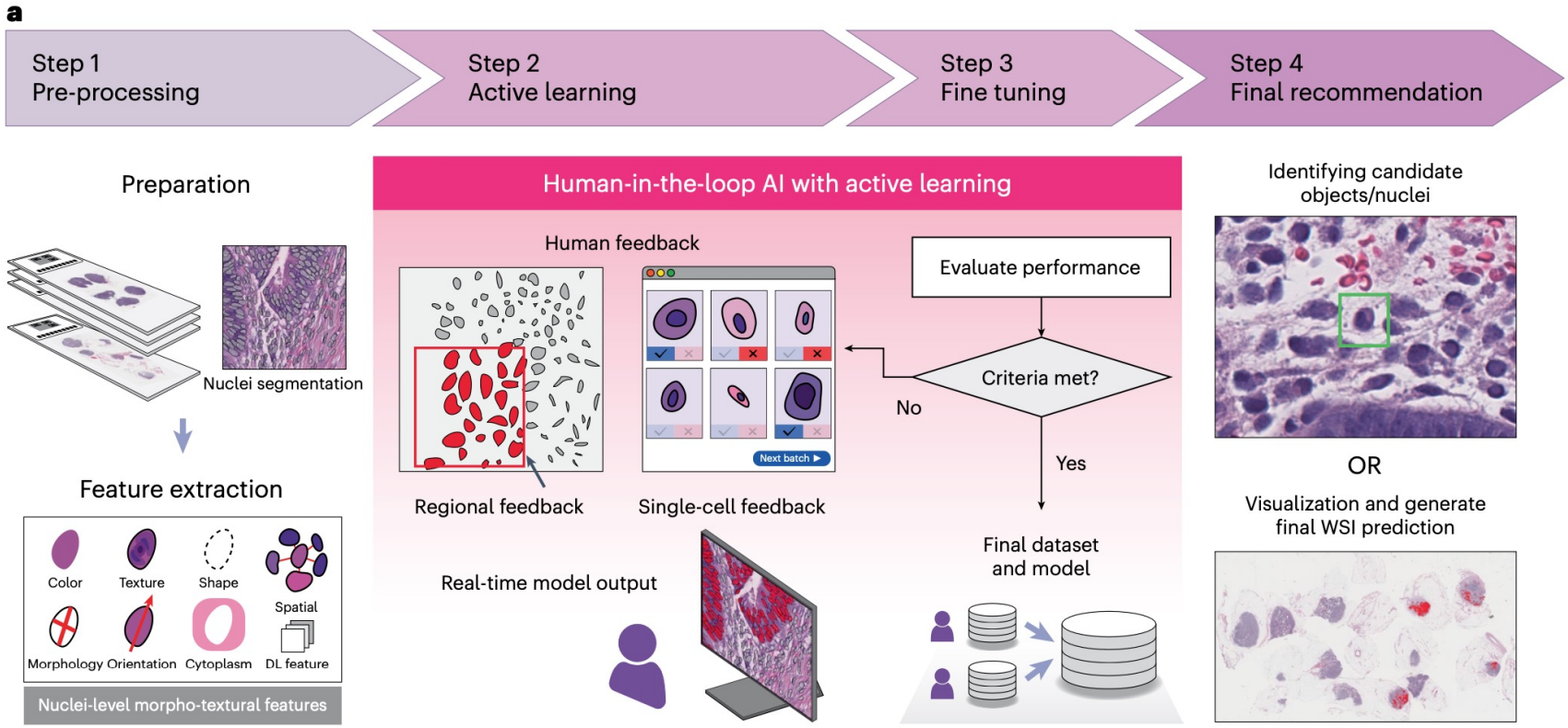
A pathologist–AI collaboration framework

- Validate the effectiveness of the framework via two crossover user studies that leveraged collaboration between the AI and the pathologist

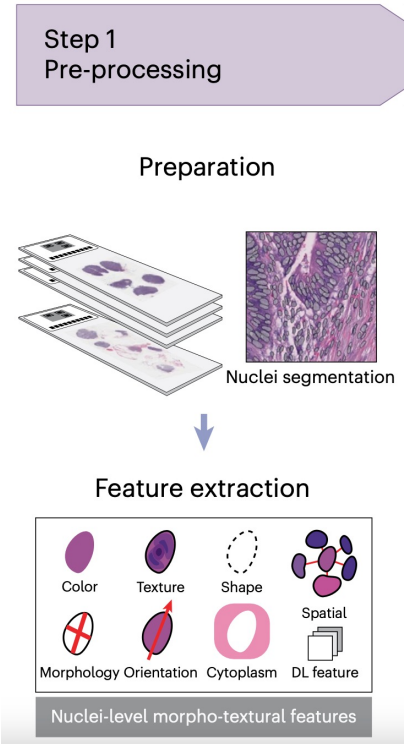
Study 1
The study specifically focuses on identifying plasma cells (PCs) in endometrial biopsies for the evaluation of chronic endometritis (CE)

Study 2:
detection of colorectal cancer metastasis in lymph nodes

Pipeline


a
Step 1
Pre-processing
Step 2
Active learning
Step 3
Fine tuning
Step 4
Final recommendation
Preparation
Nuclei segmentation
Feature extraction
Color
Texture
Shape
Spatial
Morphology
Orientation
Cytoplasm
DL feature
Nuclei-level morpho-textural features
Human-in-the-loop AI with active learning
Human feedback
Regional feedback
Single-cell feedback
Next batch
Real-time model output
Evaluate performance
Criteria met?
No
Yes
Final dataset
and model
Identifying candidate
objects/nuclei
OR
Visualization and generate
final WSI prediction

Stage 1

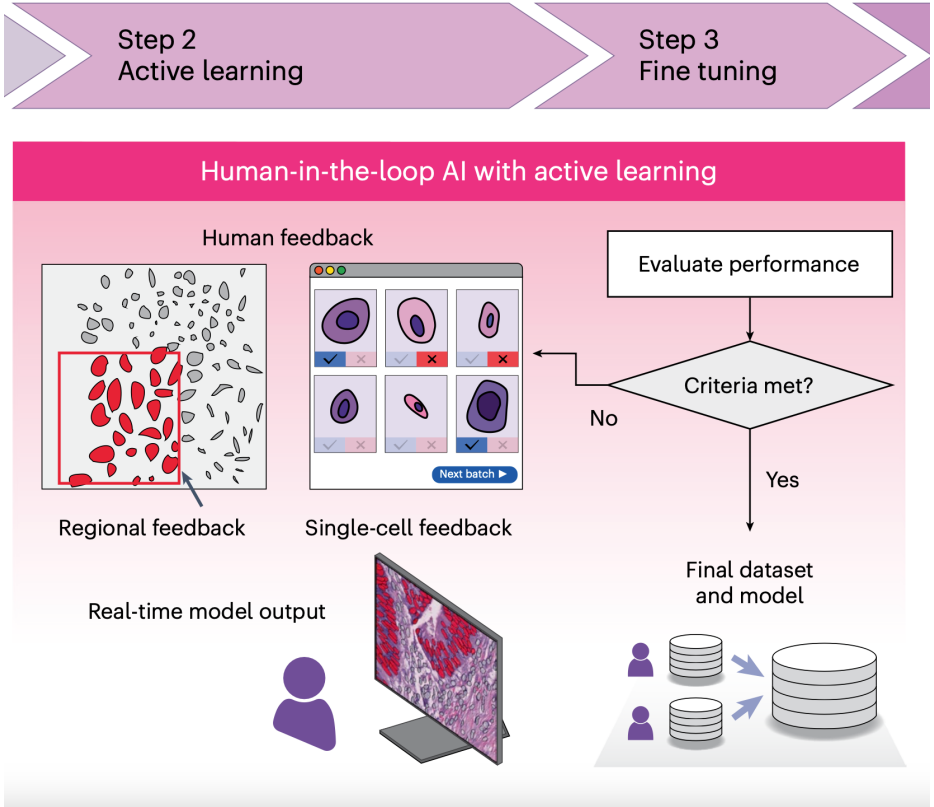


In stage 1, the slides were digitally scanned and underwent our WSI nuclei segmentation pipeline (Methods).

Next, a single-nucleus-level feature extraction pipeline was employed, using both handcrafted interpretable features and deep learning features,

Among the 154 interpretable features, 90 of them are newly introduced in this study, for example, 32 PC features that adopted a new strategy of using polar coordinates.

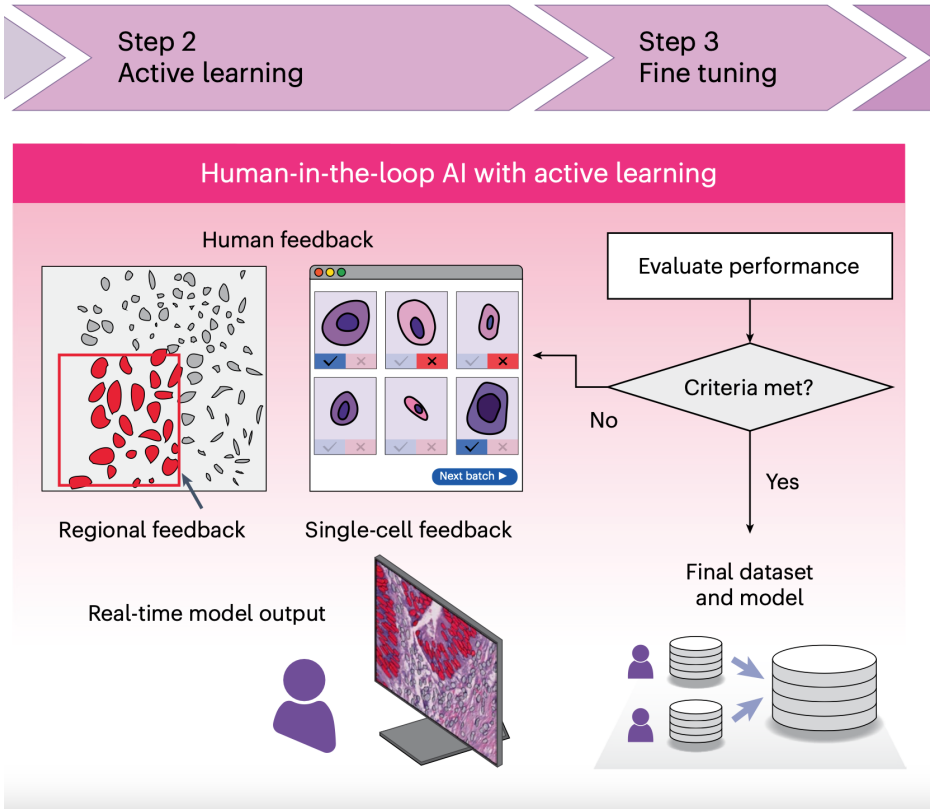
Stage 2



In the second stage of the framework, the pathologists interact with the system to provide real-time feedback by annotating nuclei of interest, such as from tumour and non-tumour cells.

The system can rapidly adapt the new data into a new ML model and populate new data points near the decision boundary for pathologists to annotate in subsequent iterations.

Stage 2

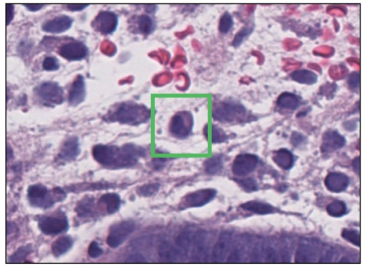


The annotation process provides two modes of user interaction: (1) regional nuclei annotation feedback and (2) single-nucleus annotation feedback, depending on how the nuclei of interest were clustered or not.

Stage 3

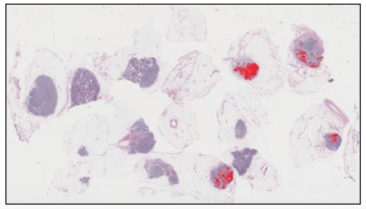
Step 4
Final recommendation

Identifying candidate objects/nuclei



OR

Visualization and generate final WSI prediction



Stage 2 iterates until the predetermined criteria are met (stage 3). The specific criteria may vary depending on the tasks and objectives.

For instance, if pathologists are satisfied with the current model performance under specific scenarios, the model may be deemed ready for deployment. The ML model can be established by a single pathologist (individualized ML model) or aggregated from multiple expert annotators into a larger dataset to train a single ML model (expert-trained ML model).

Study design with the AI system

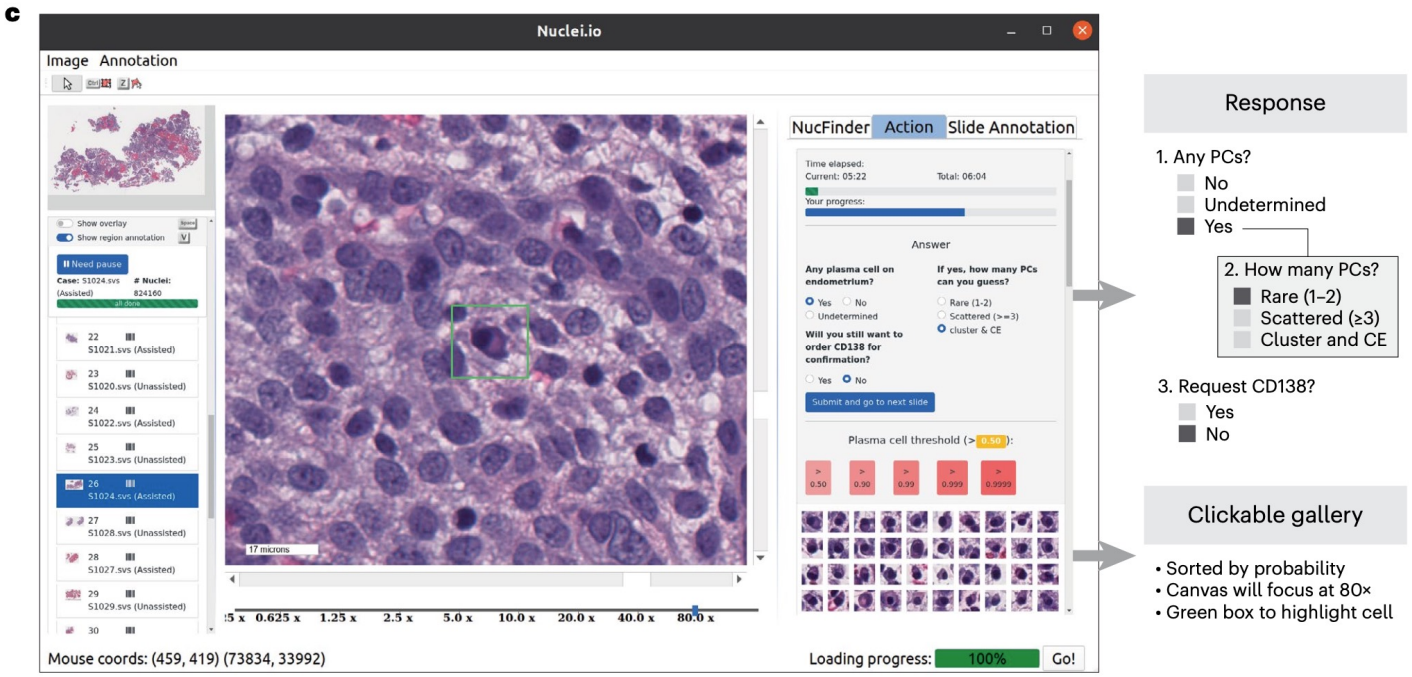
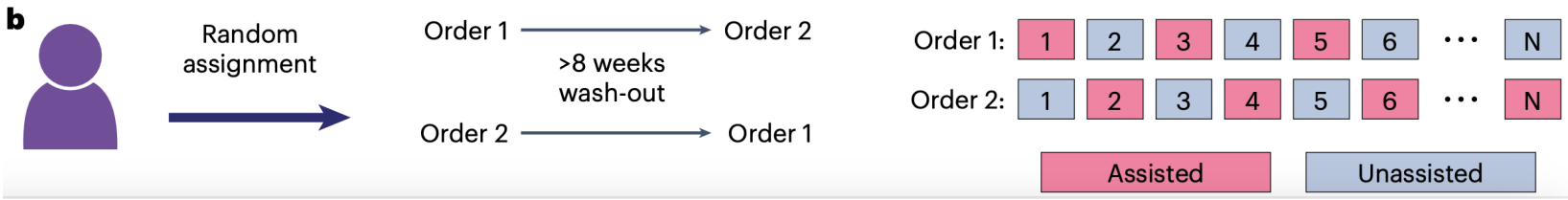


Fig. 1 | Overview of the nuclei.io framework and of the study design. a, Human-in-the-loop active learning for digital pathology from WSI pre-processing to the final recommendation. **b**, Crossover study design. The pathologists are presented with slides in alternating modes of assisted and unassisted. **c**, Screenshot of the nuclei.io interface used for PC identification.

Study design with the AI system



Order 1 began with the case in assisted mode, while order 2 started with the case in unassisted mode. Both orders alternated ML assistance during the experiment for different cases.

After 8 weeks wash-out, the participants were invited to evaluate the same cases under the same sequence but with switched order.

Identifying PCs in endometrial biopsy with expert-trained ML model

- There is currently no ML model for classifying PCs in endometrial biopsies. Locating a single PC among thousands of cells can be quite challenging, pathologists rely on CD138 IHC to highlight PCs43.

- Rather than relying on CD138 IHC, potential for ML model to assist pathologists with PC identification based solely on H&E-stained WSIs.

- After inspection, pathologists were asked to respond to three questions: '(1) Are there any PCs present in the endometrium? (2) If the answer to (1) is yes, how many PCs were found? (3) After inspection, is there a need to request CD138 IHC?'

Identifying PCs in endometrial biopsy with expert-trained ML model

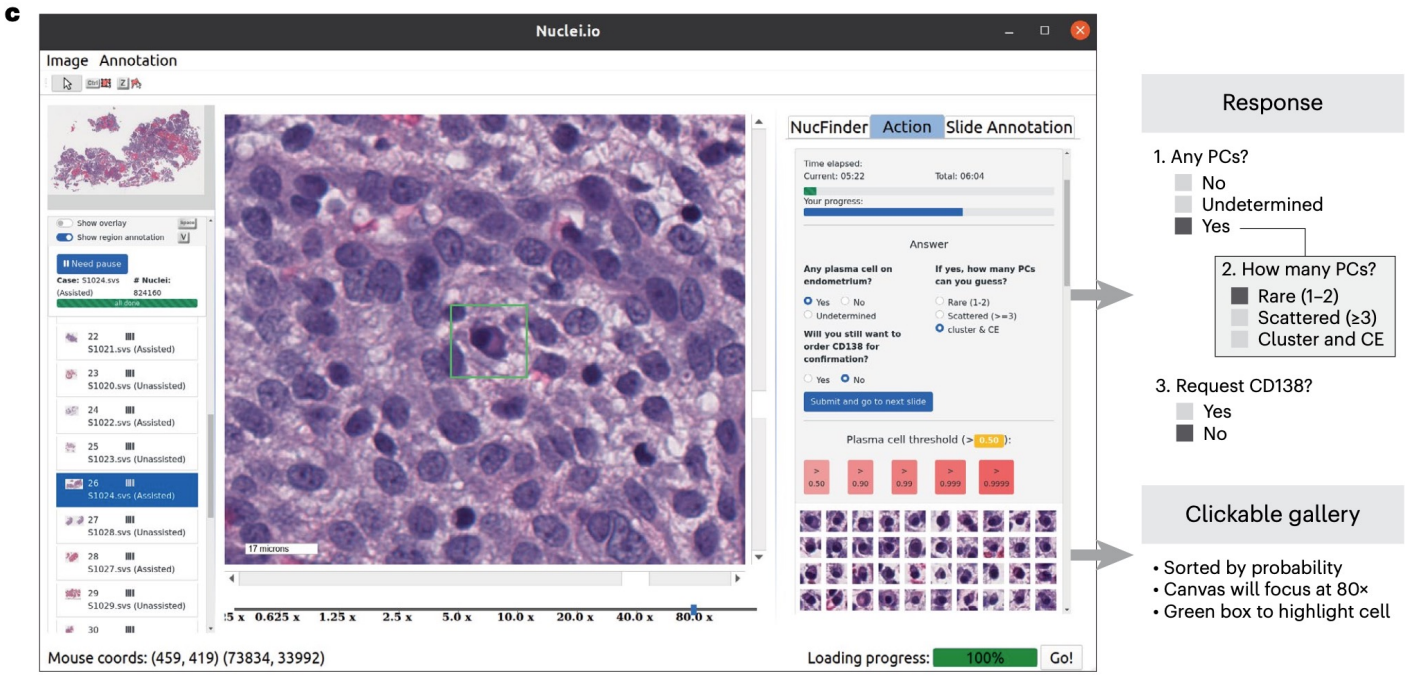


Fig. 1 | Overview of the nuclei.io framework and of the study design. a, Human-in-the-loop active learning for digital pathology from WSI pre-processing to the final recommendation. **b**, Crossover study design. The pathologists are presented with slides in alternating modes of assisted and unassisted. **c**, Screenshot of the nuclei.io interface used for PC identification.

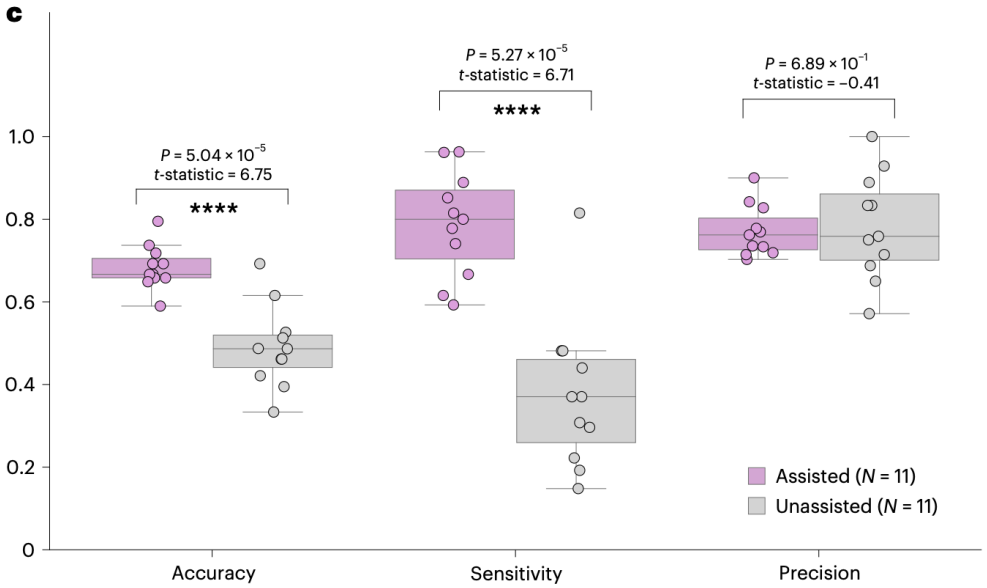
Findings

- Pathologist–AI collaboration improves PC identification
- Pathologist–AI collaboration speeds up identification of PCs
- Individualized ML model improves LN metastasis identification in CRC
- Individualized ML model speeds the detection of LN metastasis

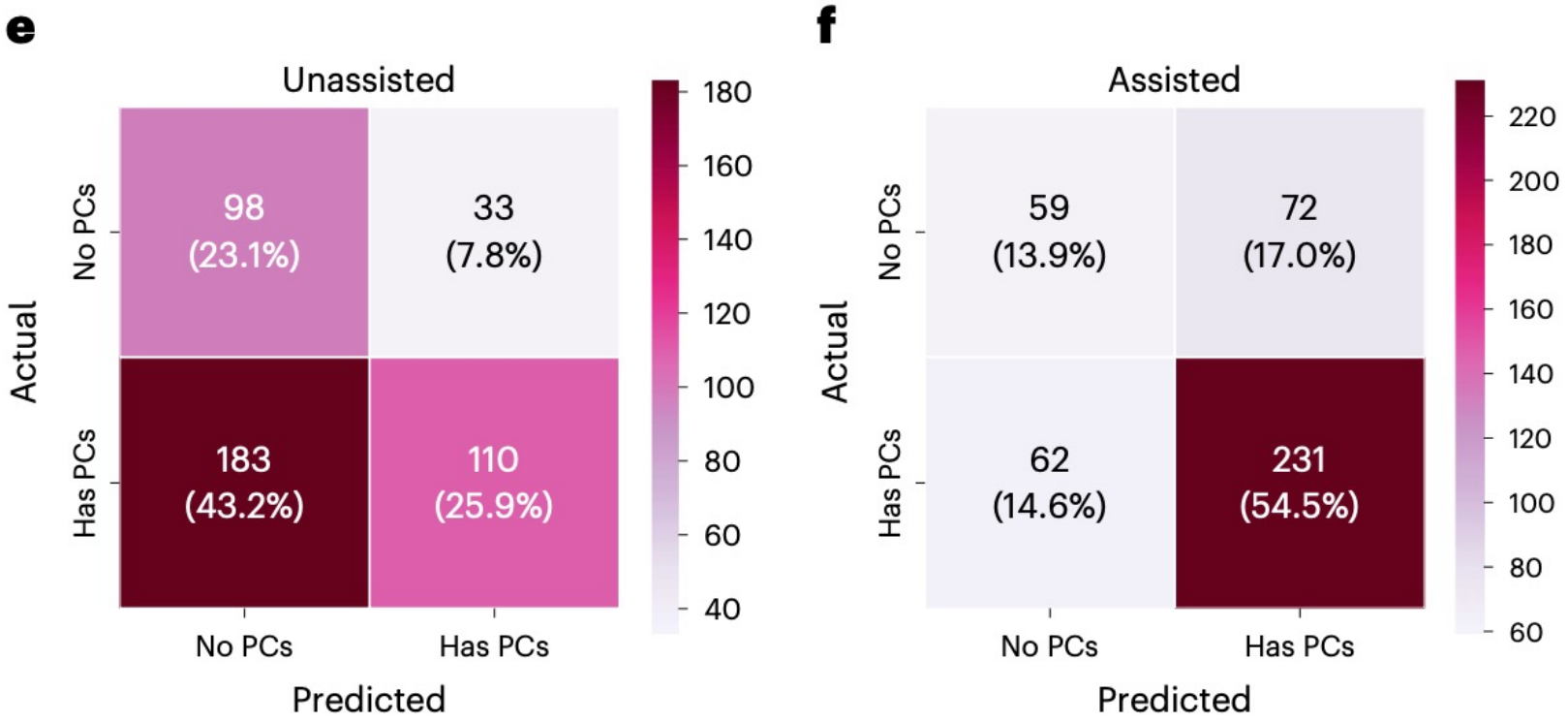
Pathologist–AI collaboration improves PC identification

The assisted mode significantly improved

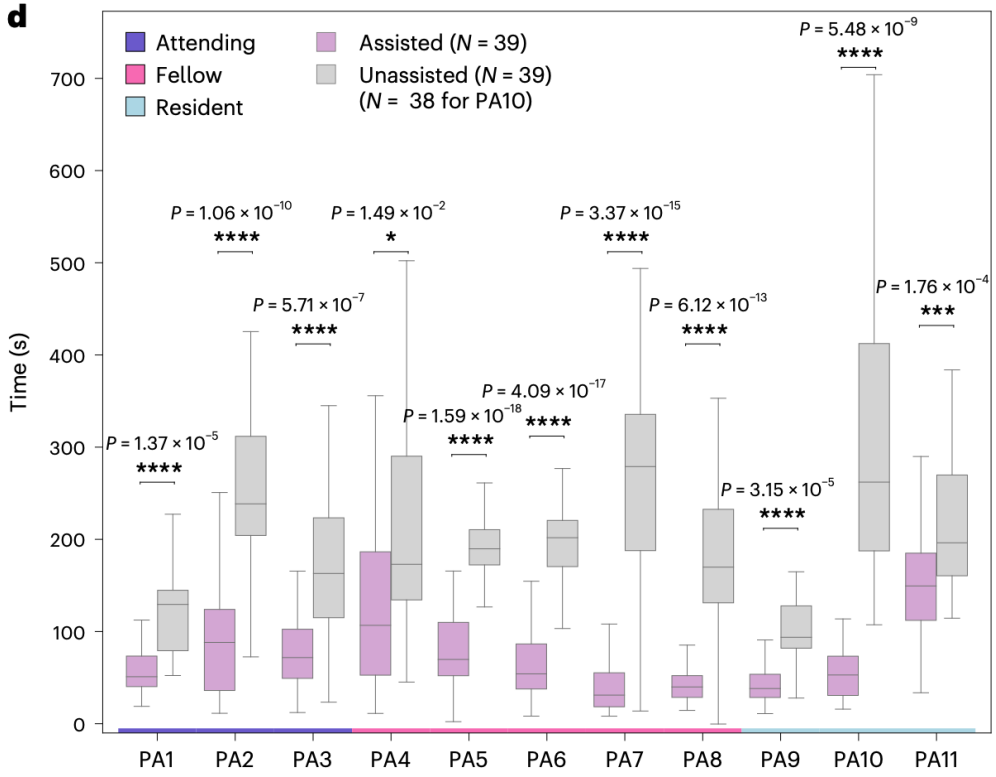
- accuracy from 0.49 (95% confidence interval (CI) 0.42–0.56) to 0.68 (95% CI 0.65–0.72) (average improvement of 0.19, P = 5.04 × 10−5)
- sensitivity from 0.38 (95% CI 0.25–0.50) to 0.79 (95% CI 0.70–0.87) (average improvement of 0.41, P = 5.27 × 10−5).
- ML assistance did not result in a significant change in precision (P = 0.689).



Pathologist–AI collaboration improves PC identification



Pathologist–AI collaboration speeds up identification of PCs



Individual time comparison between assisted mode and unassisted mode for each pathologist.

The average time for the assisted mode was 79.52 s, while the unassisted mode's average time was 209.52 s,

Individualized ML model improves LN metastasis identification in CRC

- Situations where pre-trained models are unavailable or where the dataset has different staining or quality that makes existing expert-trained models problematic. In such scenarios, it is important to enable pathologists to efficiently build personalized models.

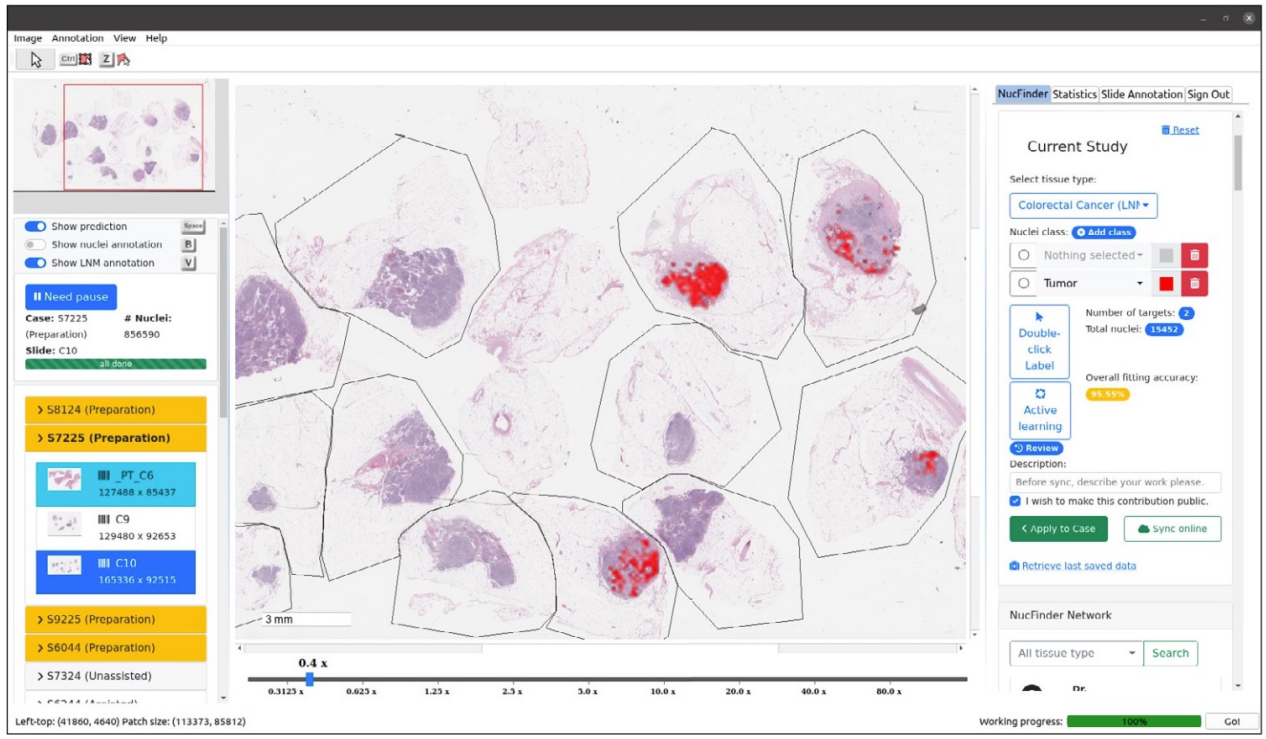
- The study sample consisted of 29 cases from Stanford Medicine in 2015, each with a single primary CRC site and at least one LN slide, which could contain multiple LNs (see Methods for details).

- Eight pathologists were invited to annotate data from the initial four cases (4 primary CRC slides and 14 LN slides) from January 2015, and the remaining 25 cases (consisting of 137 LN slides) were used for the experiments (Fig. 4a and Supplementary Table 3)

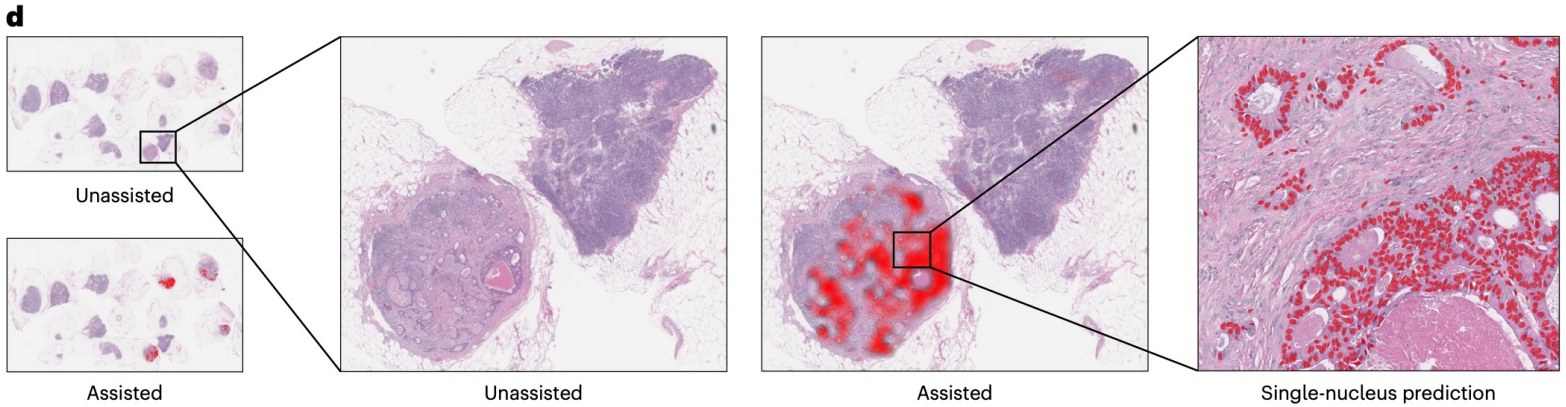
- This approach not only enabled effective active learning during the training phase but also provided accountable ML during the experimental phase, thereby helping to establish trust between pathologists and ML model.

Individualized ML model improves LN metastasis identification in CRC

c



Individualized ML model improves LN metastasis identification in CRC



Compared with unassisted WSI, the assisted mode in this study highlighted tumour regions in red colour as a heatmap on low magnification

Accuracy: from 0.9665 to 0.9694 (average improvement of 0.0028)
F1 score from 0.88 to 0.89 (average improvement of 0.01)
Sensitivity: from 0.78 to 0.80 (average improvement of 0.02).

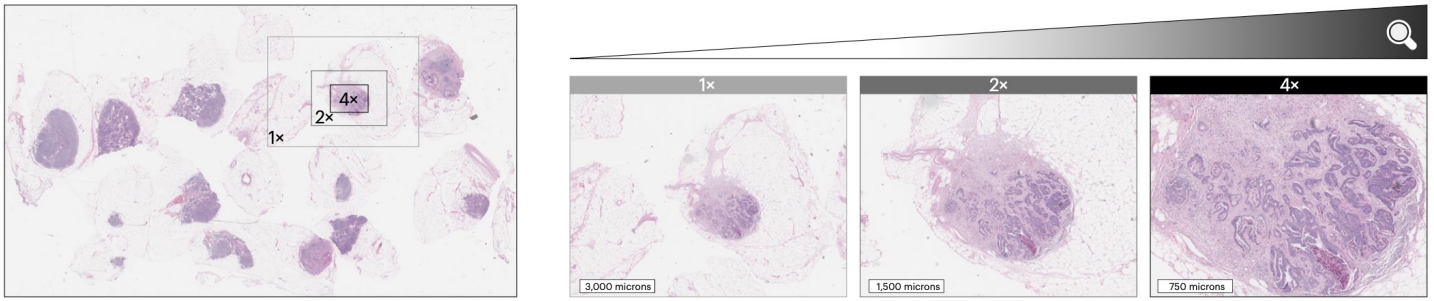
Individualized ML model improves LN metastasis identification in CRC

f

	Sensitivity			True positives		
	Assisted	Unassisted	Changes	Assisted	Unassisted	Changes
ITC	0.1979	0.1146	+0.0833	19	11	+8 (72.73%)
Micromet	0.6875	0.6731	+0.0144	143	140	+3 (2.14%)
Macromet	0.9886	0.9795	+0.0091	435	431	+4 (0.93%)
All LN+	0.8024	0.7823	+0.0202	597	582	+15 (2.58%)

Individualized ML model speeds the detection of LN metastasis

a



b

	Magnification >1×				Magnification >2×				Magnification >4×			
	Assisted	Unassisted	Changes	*P*	Assisted	Unassisted	Changes	*P*	Assisted	Unassisted	Changes	*P*
All negatives	14.04s	18.15s	−4.11s (−22.65%)	1.32×10^{-50} ****	12.33s	16.84s	−4.51s (−26.79%)	2.47×10^{-61} ****	9.00s	13.42s	−4.42s (−32.91%)	6.98×10^{-62} ****
All positives	12.30s	15.21s	−2.91s (−19.11%)	6.25×10^{-5} ****	10.48s	13.62s	−3.13s (−23.00%)	8.75×10^{-6} ****	7.51s	10.65s	−3.14s (−29.47%)	3.14×10^{-6} ****
ITC	22.52s	25.51s	−2.98s (−11.70%)	3.30×10^{-1}	20.43s	23.92s	−3.49s (−14.57%)	2.42×10^{-1}	16.38s	20.35s	−3.98s (−19.53%)	1.55×10^{-1}
Micromets	17.92s	23.15s	−5.24s (−22.61%)	1.45×10^{-3} **	15.74s	20.88s	−5.14s (−24.63%)	1.32×10^{-3} **	11.25s	16.07s	−4.82s (−30.01%)	2.11×10^{-3} **
Macromets	7.42s	9.21s	−1.79s (−19.41%)	8.45×10^{-3} **	5.83s	7.93s	−2.10s (−26.52%)	1.31×10^{-3} **	3.81s	5.96s	−2.16s (−36.18%)	4.57×10^{-4} ***

Challenges

Challenges in grasping the paper's methodology:

- No methods description
- The codebase needs debugging when running with MacOS

The pros and cons in collaborative AI research:

- Pros: Interesting topic, no need a lot of data so can deal with rare medical cases
- Cons:
    - Need team of pathologist and time, human annotation resources
    - Familiar with various interfaces, GUI programming

Thank you!