Getting Started with AI in Cytopathology

Path-iQ Getting Started Series · No. 2 of 20 · A practical primer for the practising pathologist

July 2026

1. Why cytopathology has the strongest case of any subspecialty

Cytopathology has a claim no other discipline can make: it hosted the first FDA-authorised AI medical device of any kind. PAPNET, a neural-network-based cervical smear rescreening system, was authorised in 1995 — a decade before “deep learning” entered the vocabulary. Automated cervical screening has been a regulated reality for thirty years.

The structural reasons are worth naming, because they explain why cytology AI keeps outperforming expectations. Screening is a high-volume, repetitive, attention-limited task with a well-characterised false-negative rate — routine cervical cytology carries a false-negative rate above 10% in many settings. The diagnostic unit is the cell, not the tissue architecture, which makes the computer vision problem more tractable. And the global cytotechnologist and cytopathologist shortage creates an economic pull that solid-tumour histopathology does not have to the same degree.

2. The evidence base is unusually strong

Cytopathology is the only subspecialty where AI has been tested at population scale with prospective and randomised designs.

The most instructive dataset is the AICCS system reported in *Nature Communications*. It was developed and tested across retrospective, prospective and randomised observational cohorts totalling 16,056 participants, using a two-model architecture — one detecting abnormal cells at patch level, another classifying the whole-slide image. In prospective assessment it achieved an AUC of 0.947, sensitivity 0.946, specificity 0.890. The finding that matters most, though, came from the randomised arm: cytopathologists assisted by the system showed a 13.3% improvement in sensitivity over cytopathologists working alone, with higher specificity and accuracy as well.

That is the correct framing. The question is not whether AI beats the cytopathologist. It is whether the assisted cytopathologist beats the unassisted one — and in cervical cytology, the answer is repeatedly yes.

Population-scale work supports this. A cohort study of approximately 700,000 women in Hubei province, China, found AI-assisted cytology delivered a 5.8% sensitivity increase over conventional manual reading. Separate prospective validation of an automated Bethesda-system reader reported sensitivity for detecting cervical intraepithelial neoplasia of 94.74%, exceeding senior cytologists reading manually.

3. Beyond the Pap: where else it works

Thyroid FNA. The Bethesda system’s indeterminate categories — AUS/FLUS and follicular neoplasm — are the single largest source of diagnostic uncertainty and unnecessary surgery in cytopathology. Machine learning applied to FNA smears, particularly combined with molecular testing, is being explored to reduce that indeterminate fraction. Evidence is promising but has not reached the maturity of cervical work.

Urine cytology. The Paris System's structured criteria for high-grade urothelial carcinoma make it well suited to automated nuclear-to-cytoplasmic ratio measurement and hyperchromasia assessment — both quantitative tasks humans perform inconsistently.

Effusions and serous fluids. Distinguishing reactive mesothelial cells from adenocarcinoma is a classic interobserver problem; automated morphometry plus immunocytochemistry quantification is an active research area.

Rapid on-site evaluation (ROSE). Perhaps the most underrated opportunity. AI-assisted adequacy assessment at the point of procedure could relieve a task that consumes disproportionate cytopathologist time and is frequently delegated by necessity.

4. Your first 90 days

Start with the scanner question, and take it seriously. Cytology scanning is genuinely harder than histology scanning. Cells sit at multiple focal depths, so single-plane scanning loses information a cytologist depends on. Z-stacking or extended-depth-of-field imaging is usually required, and it multiplies file sizes and scan times. Confirm your scanner supports it before you evaluate any algorithm.

Choose a screening use case, not a diagnostic one. The safest and most defensible entry point is AI-assisted primary screening or rescreening in cervical cytology, where the human retains sign-out authority and the algorithm ranks or flags. Locate-and-present workflows — where the system serves the cytotechnologist a gallery of the most abnormal objects — have the longest track record and the clearest regulatory precedent.

Baseline your false-negative rate. Use your existing rescreen and retrospective-review data. Cervical cytology laboratories already collect this for quality assurance; you likely have the denominator you need.

Shadow-run for a full quarter. Cytology case mix varies seasonally with screening programme cycles. A four-week pilot will not show you the tail.

5. Validation specifics for cytology

Preparation type is a first-class validation variable. An algorithm validated on ThinPrep will not necessarily perform on SurePath, and neither will transfer cleanly to conventional smears. Confirm the preparation the model was trained on and validate on yours.

Cellularity and adequacy handling must be tested explicitly. Ask what the system does with a hypocellular slide, an obscured slide, and a slide with heavy inflammation or blood. In screening, a wrongly "negative" inadequate slide is the most dangerous single failure.

HPV co-testing changes the arithmetic. In programmes that have moved to primary HPV screening with cytology triage, the cytology case mix becomes enriched for abnormality. An algorithm calibrated on an unselected screening population will be miscalibrated in triage. Ask the vendor which population the operating point was set on.

6. Failure modes specific to cytopathology

Focal plane loss. The commonest technical failure. Nuclear detail assessed on a single focal plane systematically under-calls hyperchromasia and chromatin abnormality.

Obscuring factors. Blood, inflammation and lubricant degrade algorithm performance more severely than human performance, because the human simply looks elsewhere on the slide while the algorithm scores what it sees.

Rare-entity blindness. Endometrial adenocarcinoma cells on a cervical preparation, herpes, and small-cell carcinoma are precisely the findings a screening-trained model may not represent well — and precisely the findings that must not be missed.

Deskilling. If the algorithm pre-selects the fields the cytotechnologist views, screening skill on full-slide review will atrophy. Departments should maintain periodic unassisted review to preserve competency and to detect algorithm drift.

7. What to watch next

Cytology is where AI is most likely to genuinely extend access rather than merely improve throughput. Deployment in settings with no cytopathologist at all — using AI as a first-pass screen with remote expert confirmation — is under active evaluation in several national programmes, and it is the application most likely to change population outcomes rather than laboratory efficiency. Watch also for foundation models fine-tuned specifically on cytological preparations; the current generation is overwhelmingly trained on tissue sections, and cytology is under-represented in their pretraining corpora.

Vendor due-diligence checklist

Use this before any procurement conversation. The first seven questions apply to every AI tool in pathology; the last three are specific to this subspecialty. A vendor unable or unwilling to answer them in writing has told you something important.

Applies to any pathology AI tool

What is the intended use statement, and what is explicitly excluded from it?

What is the regulatory status in my jurisdiction — cleared, CE-marked, research use only?

Can I see performance on an independent external cohort, not only internal validation?

What is the prevalence of the target condition in the validation cohort versus my practice?

What is the failure mode when the model is uncertain — does it abstain, or always answer?

What monitoring does the vendor provide for performance drift after deployment?

Who is liable if an algorithmic output contributes to a diagnostic error?

Specific to this subspecialty

Which preparation types was this validated on — ThinPrep, SurePath, conventional smear?

Does the system require z-stacking or extended depth of field, and does my scanner support it?

How does it behave on hypocellular, obscured and heavily inflammatory slides?

Ask for answers in writing, and keep them. When the tool underperforms in your laboratory — and at some point it will — the gap between what was claimed and what was validated is where the conversation starts.

8. Key references

#	Source	Note
1	AI enables precision diagnosis of cervical cytology grades, *Nature Communications*, 2024	AICCS, 16,056 participants; +13.3% assisted sensitivity
2	AI-assisted cytology in large-scale cervical screening, population cohort	~700,000 women; +5.8% sensitivity
3	AI in cytopathological applications for cancer, *Eur J Med Res*, 2024	Systematic review, 34 studies
4	FDA AI-Enabled Medical Device List	PAPNET, 1995 — first authorised AI device
5	The Bethesda System / The Paris System	Reporting frameworks underpinning automation

Glossary of terms used in this primer

Whole-slide image (WSI) — a digitised glass slide, typically gigapixel in scale, produced by a scanner and viewed on screen rather than under a microscope.

Foundation model — a large model pretrained on vast unlabelled image data to learn general-purpose visual features, then adapted to specific tasks with comparatively little labelled data.

Weakly supervised learning — training where only a slide-level label is available (for example “contains cancer”) rather than region-by-region annotation. Most clinical pathology AI is built this way, because exhaustive annotation is unaffordable.

AUROC — area under the receiver operating characteristic curve, a single figure summarising discrimination across all thresholds. 0.5 is chance, 1.0 is perfect. It says nothing about calibration or about performance at the threshold you would actually use.

Calibration — whether a model’s stated probability matches observed frequency. A well-calibrated model that says “80% confident” is right about 80% of the time. Calibration matters more than accuracy when a clinician acts on the number.

Abstention — the model declining to answer when confidence falls below a threshold, rather than returning its best guess. Essential wherever rare entities are in play.

Batch effect — systematic differences arising from scanner, stain lot, laboratory or site rather than from biology. The leading cause of models performing worse elsewhere than where they were built.

Automation bias — the human tendency to defer to a system that is usually correct, and so to stop checking independently. A property of the workflow, not of the algorithm.

Predetermined Change Control Plan (PCCP) — a regulator-agreed specification of changes a manufacturer may make to an authorised device without a new submission. Its existence acknowledges that these systems are not static.

Shadow deployment — running an algorithm in parallel with routine practice, without acting on its output, to characterise its behaviour on your own material before it touches a live workflow.

Path-iQ — Global Pathology AI Review. Prepared for practising pathologists beginning AI adoption.