Getting Started with AI in Pulmonary Pathology

Path-iQ Getting Started Series · No. 12 of 20 · A practical primer for the practising pathologist

July 2026

1. Two distinct territories: thoracic oncology and interstitial lung disease

Pulmonary pathology divides cleanly for AI purposes. Thoracic oncology — lung carcinoma subtyping, biomarker assessment, staging — is a high-volume, well-funded, heavily studied domain. Interstitial lung disease is a low-volume, pattern-based, multidisciplinary domain where the pathologist's contribution is one input among several.

Both benefit from AI, but for different reasons and on different timelines.

2. Thoracic oncology: where the work has concentrated

Lung adenocarcinoma versus squamous cell carcinoma classification from H&E was one of the earliest and most-cited demonstrations that deep learning could perform histological classification at expert level — and the same models were shown to predict several driver mutations directly from morphology. That result, from 2018, effectively launched the morphology-to-molecular research programme across all of pathology.

Since then the thoracic tasks have consolidated:

Histological subtyping. Adenocarcinoma versus squamous versus small cell versus large cell neuroendocrine. Clinically consequential, since treatment differs substantially.

Adenocarcinoma growth pattern quantification. Lepidic, acinar, papillary, micropapillary and solid patterns, with the predominant pattern determining grade under current grading systems. Pattern quantification is exactly the kind of tedious, area-based measurement that humans estimate and computers measure.

PD-L1 tumour proportion score. The most immediately practical application in the subspecialty. PD-L1 TPS determines immunotherapy eligibility, the scoring is a percentage, and interobserver agreement — particularly around the 1% and 50% thresholds — is a well-documented problem. Automated TPS quantification directly addresses a treatment-determining, poorly reproducible measurement.

Driver mutation inference. *EGFR*, *ALK*, *KRAS* and others inferred from H&E. Research-grade performance; a triage aid at best, never a substitute for molecular testing where a targeted therapy decision depends on it.

Spread through air spaces (STAS). Prognostically important, recently defined, and assessed inconsistently — including genuine debate about artefact versus true STAS.

Mesothelioma. Distinguishing reactive mesothelial proliferation from malignant mesothelioma is one of the hardest calls in pathology, with major medicolegal consequences. BAP1 loss and CDKN2A deletion have

improved this; algorithmic support remains exploratory.

3. Interstitial lung disease

The usual interstitial pneumonia pattern and its mimics are diagnosed by architectural pattern — patchy fibrosis, fibroblast foci, honeycombing, spatial and temporal heterogeneity. This is a whole-slide, low-magnification, gestalt task, quite different from cell-level detection.

Two things make it hard for AI. First, volume: surgical lung biopsies for ILD are uncommon and declining as transbronchial cryobiopsy and radiological diagnosis expand. Second, the diagnosis is formally multidisciplinary — the pathological pattern is interpreted alongside high-resolution CT and clinical history, and the final label emerges from discussion. A model that outputs “UIP pattern” is producing one input to that discussion, not a diagnosis.

That said, quantitative fibrosis measurement in ILD is a genuine opportunity, particularly given the precedent set in hepatology where an AI histology tool has been formally qualified for use in clinical trials. Antifibrotic drug development in the lung needs reproducible histological endpoints for the same reasons.

4. Your first 90 days

Weeks 1–2: start with PD-L1. It is the clearest case in the subspecialty. Audit your current interobserver agreement at the 1% and 50% thresholds — the two cut-points that determine therapy — and you will have your business case.

Weeks 3–6: evaluate a PD-L1 quantification module in shadow mode. Compare automated TPS against your manual score on consecutive cases. Log every case where the automated and manual scores fell on opposite sides of a treatment threshold.

Weeks 7–12: extend to adenocarcinoma pattern quantification if you resect meaningful volumes. This is measurable, retrospectively verifiable against archived cases, and feeds directly into grading.

5. Validation specifics

Small biopsy and cytology material dominates real thoracic practice. Most lung cancers are diagnosed on bronchoscopic biopsy, EBUS-guided fine-needle aspiration or core needle biopsy — not resection. Models trained on resections systematically overestimate their applicability. Confirm the specimen type.

PD-L1 assay identity is critical. Different clones and platforms are validated for different therapies with different cut-points and are not interchangeable. An algorithm must be validated for the specific assay you run.

Anthracosis, haemorrhage and crush artefact are routine in lung specimens and degrade image analysis disproportionately.

Prior treatment, including immunotherapy, produces marked histological change — necrosis, fibrosis, dense lymphoid infiltrates and cholesterol clefts — that is under-represented in training data derived from treatment-naive resections.

6. Failure modes to anticipate

Reactive pneumocyte atypia. Organising pneumonia, diffuse alveolar damage and post-treatment change produce atypia that mimics malignancy. This is a leading cause of false positives.

Neuroendocrine differentiation. Small cell carcinoma on a crushed biopsy is among the most difficult diagnoses in the field for both humans and algorithms.

PD-L1 compartment errors. TPS counts tumour cells only. Immune cell staining and macrophage staining are common sources of automated over-scoring, and this is the specific failure to test for during shadow deployment.

Mesothelial proliferation. The reactive-versus-malignant question should not be delegated to an image model on current evidence.

7. What to watch next

Quantitative histological endpoints for pulmonary fibrosis trials are the development most likely to change practice, following the regulatory template established in liver disease. In oncology, expect continued movement toward H&E-based biomarker inference as a triage layer that decides which molecular tests to order first — a workflow gain rather than a diagnostic one, but a real one given the tissue-scarcity problem that defines thoracic pathology.

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 PD-L1 assay, clone and cut-point is this validated for?

Was it trained on resections or on small biopsy and cytology material?

How does it exclude immune cell and macrophage staining from tumour proportion score?

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

Coudray et al., *Nature Medicine*, 2018

Lung subtype and mutation prediction from H&E

2

AI and digital tools in cancer pathology, *Lancet Digital Health*, 2025

PD-L1 CPS/TPS image analysis

3

Prov-GigaPath evaluation

Highest ranking on lung cancer tasks in benchmark study

4

FDA qualification of AI histology tool for MASH trials, 2025

Regulatory template for fibrosis endpoints

5

WHO Classification of Thoracic Tumours, 5th edition

Grading and pattern definitions

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.