Getting Started with AI in Cardiovascular Pathology

Path-iQ Getting Started Series · No. 13 of 20 · A practical primer for the practising pathologist

July 2026

1. A small subspecialty with one exceptionally well-suited task

Cardiovascular pathology is low-volume in most departments, and much of its work is autopsy-based or highly specialised. But it contains one application that is close to ideal for machine learning: grading of cardiac allograft rejection on endomyocardial biopsy.

The reasons are worth setting out, because they illustrate what makes a good AI target generally. The specimen is standardised — a handful of small myocardial fragments obtained on a fixed surveillance schedule. The grading system is defined and internationally agreed. The volume, in a transplant centre, is substantial and recurring. And critically, interobserver agreement on rejection grading is known to be poor, particularly at the boundary between grade 1R and 2R that determines whether a patient receives augmented immunosuppression.

Deep learning applied to endomyocardial biopsy assessment of allograft rejection has been developed and reported in the major clinical literature, with models trained across institutions and evaluated against expert consensus grading. This is the most developed application in the subspecialty and the sensible place to begin any local project.

2. The task map

Cardiac allograft rejection grading. Acute cellular rejection grading per ISHLT criteria, and detection of antibody-mediated rejection features. The highest-value application.

Myocarditis assessment. Inflammatory cell quantification and myocyte damage assessment — a diagnosis with notoriously poor reproducibility and significant clinical consequences, particularly in the era of immune checkpoint inhibitor myocarditis, which has made this a far more common question.

Fibrosis quantification. Interstitial and replacement fibrosis measurement on trichrome or picrosirius red, relevant to cardiomyopathy assessment, prognosis and increasingly to trial endpoints.

Amyloid detection and typing. Cardiac amyloidosis is substantially under-diagnosed, and its recognition now carries direct therapeutic consequence given available ATTR-directed therapy. Automated detection of Congo red birefringence and quantification of deposit burden is a well-defined image analysis task.

Atherosclerotic plaque characterisation. Largely a research application, quantifying necrotic core, fibrous cap thickness, calcification and inflammatory infiltrate — with relevance to cardiovascular device and drug development.

Sudden cardiac death investigation. Overlaps with forensic pathology; conduction system examination and structural assessment in the young sudden death case.

3. Your first 90 days

Weeks 1–3: assess whether you have the volume. This is the honest first question in cardiovascular pathology. Outside a transplant centre or a specialist cardiac unit, you may not have enough cases to validate anything locally. If that is your situation, the right strategy is participation in a multi-centre validation rather than a solo pilot — and that is a legitimate and valuable form of engagement.

Weeks 4–8: if you are a transplant centre, baseline your rejection grading agreement. Circulate archived endomyocardial biopsies among your reporting pathologists and measure agreement at the 1R/2R boundary specifically. This is the number that matters.

Weeks 9–12: consider fibrosis quantification as a parallel, low-risk project. It requires no diagnostic decision, uses stains you already perform, and can be validated against archived cases with echocardiographic or MRI correlates.

4. Validation specifics

Specimen fragmentation is intrinsic. Endomyocardial biopsies arrive as several small, often distorted fragments with variable orientation and frequent crush artefact. Any model must be validated on realistic material, not on curated well-oriented fragments.

Biopsy site artefact is a specific and important confounder. Previous biopsy sites show granulation tissue, fibrosis and inflammatory infiltrate that closely mimic rejection — the classic “Quilty” and healing biopsy site problems. These must be represented in training and explicitly tested.

Levels and step sections matter. Rejection assessment depends on examining multiple levels; a model reading a single level assesses less tissue than the pathologist does.

Autopsy tissue differs from surgical tissue. Autolysis, post-mortem interval and fixation differences make autopsy cardiac tissue a distinct domain. Do not assume transfer.

5. Failure modes to anticipate

Quilty lesions. Endocardial lymphoid aggregates that mimic rejection when they extend into myocardium, and a well-known source of over-grading by humans and algorithms alike.

Ischaemic injury. Peri-transplant ischaemic damage in early post-transplant biopsies resembles rejection-associated myocyte injury.

Infection. Cytomegalovirus and toxoplasma infection produce inflammatory infiltrates that a rejection-trained model will grade as rejection.

Antibody-mediated rejection. Diagnosis depends on immunohistochemistry for C4d and CD68 alongside donor-specific antibody results — a multimodal problem that an H&E model cannot address.

Low prevalence of high grades. Severe rejection is rare in modern immunosuppression eras, so training data is heavily imbalanced toward grade 0 and 1R. Per-class sensitivity for higher grades is the number to demand.

6. What to watch next

Two directions. First, the extension of quantitative fibrosis endpoints into cardiac clinical trials, following the regulatory precedent set in liver disease — reproducible histological fibrosis measurement is a plausible surrogate endpoint and the methodology now exists. Second, integration of endomyocardial biopsy assessment with donor-derived cell-free DNA and gene expression profiling, both of which are already used to reduce surveillance biopsy frequency. The interesting question for the next few years is not whether AI can grade a biopsy, but whether the combination of molecular surveillance and AI-assisted histology safely reduces the number of biopsies performed at all.

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

Was this validated on fragmented, crush-artefacted endomyocardial biopsy material?

How does it handle Quilty lesions, previous biopsy sites and peri-transplant ischaemic injury?

What is per-class sensitivity for higher rejection grades, which are rare in training data?

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.

7. Key references

#

Source

Note

1

ISHLT working formulation for cardiac allograft rejection grading

Grading framework underpinning automation

2

Deep learning assessment of cardiac allograft rejection from endomyocardial biopsies, *Nature Medicine*, 2022

Principal published application *[verify citation details before publication]*

3

AI-assisted fibrosis scoring precedent, *JHEP Reports*, 2026

Cross-organ fibrosis quantification template

4

Benchmarking foundation models, *Nature Biomedical Engineering*, 2025

Transportability considerations

Note: cardiovascular AI literature is smaller than in oncological subspecialties. Verify current tool availability and citation details locally before procurement or publication.

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.