COMPUTATIONAL PATHOLOGY ENGINEERING

From Histology Patches and CNNs to Whole-Slide Foundation Models, Explainable AI, and Production MLOps

Faramarz Kowsari

Book DOI	10.5281/zenodo.21444837
DOI URL	https://doi.org/10.5281/zenodo.21444837
Google Books	GGKEY:8ZWNQ7NFGBL
Edition	Sample Placeholder Edition

SAMPLE PLACEHOLDER EDITION

This temporary PDF establishes stable repository links, metadata, citation identifiers, and the planned teaching structure. It contains no clinical performance claim and will be replaced by the final infographic edition using the same filename.

1. Purpose and Project Relationship

This book is the academic and teaching layer of the Computational Pathology AI Lab. The repository demonstrates reproducible code, tests, configuration, deployment, and evidence controls; the book explains why those engineering choices matter in medical image analysis.

The project originated as a graduate deep-learning assignment at the University of Colorado Boulder. Its modernization preserves that history while separating original evidence from new research software.

Evidence rule

Executed evidence must be traceable to data provenance, a split manifest, configuration, logs, predictions, checksums, and a versioned commit. Planned experiments are labeled Not Yet Benchmarked. No placeholder number is presented as scientific performance.

2. Proposed Book Architecture

Part I - Foundations	Digital pathology, histology patches, WSI structure, labels, leakage, provenance, and reproducible experiment design.
Part II - Deep Learning Baselines	CNN construction, transfer learning, augmentation, stain variation, optimization, and evaluation beyond accuracy.
Part III - Modern Representations	Vision Transformers, self-supervised learning, foundation-model adapters, linear probing, and controlled fine-tuning.
Part IV - Whole-Slide Intelligence	Tissue masking, patch sampling, feature extraction, attention MIL, slide heatmaps, and computational tradeoffs.
Part V - Trustworthy Medical AI	Calibration, uncertainty, false-negative analysis, explainability limits, bias, domain shift, and external validation.
Part VI - Production Engineering	Packaging, tests, DVC, MLflow, model registry, APIs, containers, monitoring, governance, and releases.

3. From Patch to Evidence

A histology classifier is not only a neural network. It is a chain of assumptions: image provenance, label quality, patient-level split, patch sampling, preprocessing, architecture, loss, threshold, calibration, and deployment context. A failure in any link can create an impressive but misleading score.

Reference pipeline

Input	Control	Output
Patch / WSI	Validation and leakage audit	Versioned manifest
Images	Stain and geometric preprocessing	Model-ready tensors
Model	Training, tracking, and checksums	Checkpoint and logs
Predictions	Metrics, calibration, intervals	Evidence report
Release	Cards, CI, API, monitoring	Research software artifact

The repository implements the pipeline with a typed Python package, YAML configurations, synthetic fixtures, unit tests, a FastAPI service, Docker assets, DVC stages, MLflow hooks, and evidence templates.

4. Model Families

Baseline CNN

Transparent reference architecture and CPU smoke tests.

Transfer Learning

ResNet, DenseNet, EfficientNet, and ConvNeXt through torchvision.

Vision Transformer

A small educational ViT plus a boundary for larger encoders.

Foundation Adapter

External legally obtained pathology encoders; weights are not redistributed.

Attention MIL

Slide-level aggregation from variable-size bags of patch features.

The book will compare these families under identical split rules, label budgets, calibration analysis, domain shifts, resource constraints, and uncertainty reporting. Any real-data result must be inserted only after execution.

5. Trustworthy Evaluation

Medical AI evaluation must describe discrimination, decision thresholds, confidence, failure modes, and transportability. The planned report includes ROC-AUC, PR-AUC, sensitivity, specificity, F1, balanced accuracy, MCC, Brier score, expected calibration error, and bootstrap confidence intervals.

Explainability and uncertainty

Grad-CAM, integrated gradients, occlusion sensitivity, attention maps, and MC dropout are implemented as model-behavior probes. They do not prove causality, biological validity, or clinical safety. Explanations must be stress-tested across seeds, shifts, and architecture families.

False negatives receive a dedicated audit because aggregate metrics can conceal clinically important miss patterns. The repository therefore produces sortable error tables and supports uncertainty-enriched review queues.

6. Whole-Slide Learning and Production MLOps

Whole-slide images are gigapixel objects. The system separates thumbnail inspection, tissue masking, patch extraction, feature encoding, and MIL aggregation. Optional OpenSlide support is isolated so the core tests remain portable.

The operational layer includes CI, Docker, DVC stages, optional MLflow tracking, a model-registry metadata structure, FastAPI endpoints, a Streamlit demonstration, and drift utilities. The default API demo predictor is visibly labeled as an untrained placeholder unless a trusted checkpoint is configured.

A production release should bind the code commit, environment, data manifest, checkpoint checksum, metrics, model card, data card, and deployment image into one reviewable evidence package.

7. Replacement and Publication Plan

The final book will be produced as a page-by-page infographic publication. The future workflow is:

1	Finalize chapter and page map.
2	Create an English Word document containing the image-generation prompt for every infographic page.
3	Generate and review individual infographic pages.
4	Assemble the final PDF and verify all rendered pages.
5	Replace this file while preserving the filename computational-pathology-engineering.pdf.
6	Update the Zenodo record associated with DOI 10.5281/zenodo.21444837 and keep Google Books key GGKEY:8ZWNQ7NFGBL consistent.

This placeholder establishes identifiers and structure; it is not the final infographic book.