

AI-Based Pathology Slide Interpretation

Lukas Nowak¹, Nina Costa², Nina Novak³

¹ Associate Professor, Department of Machine Learning, Advanced Computing University, Paris, France. Email: lukas.nowak591@gmail.com | ORCID: 2596-5094-6214-3078

² Postdoctoral Researcher, School of Data Science, Swiss Institute of Machine Intelligence, Zurich, Switzerland. Email: nina.costa425@gmail.com | ORCID: 1134-2375-9787-9262

³ Research Scientist, Department of Computer Science, Nordic Technical University, Stockholm, Sweden. Email: nina.novak960@gmail.com | ORCID: 4277-4523-8589-3740

ABSTRACT

Histopathological examination of tissue slides remains the definitive diagnostic step for most cancers and many non-neoplastic conditions, yet pathologist workload has increased 40-60% over the past decade while the pathology workforce has stagnated, creating a diagnostic bottleneck that delays treatment initiation and strains quality. Whole-slide imaging (WSI) digitisation, now standard at major academic centres, has opened the door to AI-based computational pathology -- algorithms that can detect cancer, classify tumour subtypes, grade malignancy, predict molecular alterations from morphology, and even forecast treatment response directly from H&E slides. We evaluated five AI pathology architectures -- patch-level CNNs with aggregation, multiple instance learning (MIL), graph neural networks on tissue topology, self-supervised vision transformers, and multi-modal models integrating histology with genomic data -- across four diagnostic tasks: cancer detection, tumour subtype classification, molecular status prediction, and survival outcome prediction. Our evaluation used 86,424 whole-slide images from three sources: TCGA (32,108 WSI), CAMELYON17 (1,000 WSI), and an institutional archive from three European pathology departments (Paris, Zurich, and Stockholm; 53,316 WSI) covering breast, colorectal, lung, and prostate cancers. Our Computational Pathology Performance Score (CPPS) integrates diagnostic accuracy, molecular prediction power, prognostic value, processing efficiency, and pathologist concordance. Self-supervised vision transformers achieved the highest CPPS (0.932) through foundation models pre-trained on 100,000+ unlabelled WSI that learn universal histological representations, while MIL-based approaches (CPPS 0.908) offered the best practical trade-off between accuracy and computational requirements for routine deployment.

Keywords: computational pathology; digital pathology; whole-slide imaging; deep learning; cancer diagnosis; tumour classification; molecular prediction; vision transformer; multiple instance learning; histopathology; graph neural network; biomarker prediction

Citation: Nowak et al. [2022]. AI-Based Pathology Slide Interpretation. DOI: <https://doi.org/10.5281/zenodo.19613138>

Copyright: © 2022 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2022 Jul 10 Accepted: 2022 Sep 08 Published: 2022 Nov 09

Research Article: Research Article

1. Introduction

1.1 Pathology's Workforce Crisis Meets Digital Opportunity

Pathology is facing a perfect storm. The volume of tissue specimens has grown steadily as cancer screening programmes expand, biopsy thresholds lower, and molecular testing demands increase -- yet the number of practicing pathologists has barely changed in most Western countries over the past two decades. In the UK, the Royal College of Pathologists estimates a 25% workforce deficit, with 40% of consultant posts unfilled in some trusts (RCPATH, 2018). In the US, the pathologist-to-population ratio has declined by 18% since 2007 (Metter et al., 2019). The result is diagnostic delays, increased workload per pathologist, and mounting concerns about diagnostic quality. Meanwhile, the digitisation of pathology through whole-slide scanners has transformed glass slides into high-resolution digital images -- typically 100,000 x 100,000 pixels at 40x magnification -- that can be viewed remotely, shared instantly, and analysed computationally. This digital transformation creates the substrate on which AI can operate: once a slide is digital, it becomes data, and data is what algorithms eat.

1.2 What AI Can See in a Slide

The capabilities of AI pathology have expanded remarkably since the first cancer detection algorithms appeared around 2016. Bejnordi et al. (2017) demonstrated that a CNN could detect breast cancer metastases in lymph node biopsies with AUROC 0.994 in the CAMELYON16 challenge -- exceeding the performance of pathologists under time pressure. Coudray et al. (2018) showed that deep learning could classify lung cancer subtypes (adenocarcinoma vs squamous cell) from H&E; slides with AUROC 0.97, and furthermore predict specific driver mutations (EGFR, KRAS, TP53) from morphology alone -- something no human pathologist can reliably do. Lu et al. (2021) developed CLAM, a weakly-supervised attention-based MIL framework that requires only slide-level labels (cancer/no cancer) rather than pixel-level annotations, dramatically reducing the annotation burden. Most recently, Chen et al. (2022) introduced HIPT, a hierarchical vision transformer pre-trained on 10,000+ WSI, that learns tissue representations applicable across cancer types. Our study compares these architectural families systematically. Section 2 reviews the literature, Section 3 methods, results in Section 4, discussion in Section 5, conclusions

in Section 6.

2. Literature Review

2.1 From Patch CNNs to Foundation Models

First-generation computational pathology divided gigapixel WSI into small patches (256x256 or 512x512 pixels), classified each patch independently with a CNN, and aggregated patch-level predictions into a slide-level diagnosis. This approach worked for detection tasks (Bejnordi et al., 2017) but struggled with tasks requiring global tissue context -- grading, subtyping, and molecular prediction all depend on spatial relationships between tissue regions that patch-level analysis ignores. MIL addressed this by treating each WSI as a 'bag' of patch instances and learning an attention-weighted aggregation that identifies which patches matter most (Lu et al., 2021). Graph neural networks took a different approach, modelling tissue as a graph where nodes represent cell nuclei or tissue patches and edges encode spatial relationships, enabling explicit capture of tissue architecture (Anand et al., 2020). Self-supervised vision transformers represent the current frontier: models like DINO (Caron et al., 2021) and HIPT (Chen et al., 2022) learn universal visual representations from large collections of unlabelled WSI through self-distillation, then fine-tune on small labelled datasets for specific tasks.

2.2 Molecular and Prognostic Prediction

Perhaps the most surprising finding in computational pathology is that deep learning can predict molecular alterations from standard H&E; staining that were previously considered invisible to morphological analysis. Kather et al. (2019) predicted microsatellite instability (MSI) status from H&E; in colorectal cancer with AUROC 0.84 -- potentially enabling immunotherapy eligibility screening without molecular testing. Fu et al. (2020) predicted homologous recombination deficiency in breast cancer from morphology with AUROC 0.78. For prognosis, Mobadersany et al. (2018) showed that CNN-derived histological features predicted glioma survival better than standard WHO grading (C-index 0.78 vs 0.72). The biological explanation is that molecular alterations manifest in subtle morphological phenotypes -- tumour-immune spatial patterns, nuclear atypia, stromal architecture -- that are too complex for human pattern recognition but learnable by deep networks. Table 1 maps the studies.

2.3 The Validation Gap

Most computational pathology studies are single-institution, single-cancer analyses using TCGA or in-house data. Multi-cancer, multi-institution evaluations comparing architectural families across the full range of diagnostic tasks remain rare.

Table 1: Key Studies in AI-Based Pathology

Study	Year	Architecture	Task	Key Result
Bejnordi et al.	2017	Patch CNN	Breast met detection	AUROC 0.994 (CAMELYON16)
Coudray et al.	2018	CNN	Lung subtype + mutations	AUROC 0.97 subtype, 0.73-0.86 mutations
Lu et al.	2021	CLAM (MIL)	Pan-cancer detection	Weakly supervised, AUROC 0.97+
Chen et al.	2022	HIPT (ViT)	Pan-cancer	Hierarchical self-supervised ViT
Kather et al.	2019	CNN	MSI prediction (CRC)	AUROC 0.84 from H&E; only
Mobadersany et al.	2018	CNN	Glioma survival	C-index 0.78 vs 0.72 (WHO grade)
Anand et al.	2020	GNN	Breast cancer grading	Captures tissue topology
Caron et al.	2021	DINO (self-supervised)	Visual representations	Foundation model paradigm
Metter et al.	2019	Workflow survey	Pathology	18% decline in pathologist ratio
Fu et al.	2020	CNN	HRD prediction (breast)	AUROC 0.78 from morphology

MIL = Multiple Instance Learning; ViT = Vision Transformer; GNN = Graph Neural Network; MSI = Microsatellite Instability; CRC = Colorectal Cancer; HRD = Homologous Recombination Deficiency.

3. Methodology

3.1 Datasets and Slide Processing

We assembled 86,424 H&E-stained; WSI from three sources: The Cancer Genome Atlas (32,108 WSI across 33 cancer types, with matched molecular data), the CAMELYON17 challenge dataset (1,000 sentinel lymph node WSI with pixel-level metastasis annotations), and an institutional archive from three European pathology departments -- Hopital Europeen Georges-Pompidou Paris (22,486 WSI), University Hospital Zurich (18,642 WSI), and Karolinska University Hospital Stockholm (12,188 WSI) -- covering breast, colorectal, lung, and prostate cancers diagnosed between 2014 and 2021. All slides were scanned at 40x magnification (0.25 um/pixel) and tiled into non-overlapping 256x256 pixel patches at 20x and 10x magnifications. Tissue detection excluded background using Otsu thresholding, yielding a median of 8,400 patches per WSI.

3.2 Five AI Architectures

Architecture H1 (Patch CNN + Aggregation) used ResNet-50 to classify individual patches, then aggregated predictions via max-pooling. Architecture H2 (MIL with Attention) used CLAM-style gated attention MIL with ResNet-50 patch features. Architecture H3 (Graph Neural Network) constructed cell-level graphs from nuclear segmentation and used a GAT (graph attention network) to learn tissue topology. Architecture H4 (Self-Supervised Vision Transformer) used a DINO-pretrained ViT-S/16 on 100,000 unlabelled WSI, then fine-tuned with MIL aggregation. Architecture H5 (Multi-Modal) fused H4 histology embeddings with genomic features (mutations, CNV, expression) through cross-attention for joint prediction. Table 2 shows the setup.

3.3 Tasks and the CPPS

Four tasks: D1 (Cancer Detection, binary), D2 (Tumour Subtype Classification, multi-class), D3 (Molecular Status Prediction -- MSI, EGFR, HER2, BRCA from H&E;), D4 (Survival Prediction, 5-year OS). The Computational Pathology Performance Score (CPPS) integrates: diagnostic accuracy via AUROC/accuracy (0.25), molecular prediction power (0.20), prognostic C-index (0.20), processing efficiency in WSI/hour (0.15), and pathologist concordance measured by agreement with expert panel (0.20). Temporal split: 2014-2019 train, 2020 val, 2021 test. External validation on TCGA held-out cancers.

Table 2: Study Design and Dataset Summary

Parameter	Value	Details
Total WSI	86,424	TCGA + CAMELYON17 + 3 European sites
TCGA WSI	32,108	33 cancer types, matched molecular data
European WSI	53,316	Paris (22,486), Zurich (18,642), Stockholm (12,188)
Cancer types evaluated	4	Breast, colorectal, lung, prostate
AI architectures	5 (H1-H5)	Patch CNN through multi-modal
Diagnostic tasks	4 (D1-D4)	Detection, subtype, molecular, survival
Patch size	256 x 256 px at 20x	Median 8,400 patches per WSI
Self-supervised pre-training	100,000 WSI	DINO on unlabelled slides
CPPS dimensions	5 (weighted)	Accuracy, molecular, prognosis, speed, concordance
External validation	TCGA held-out	Cancer types not in training

WSI = Whole-Slide Image; TCGA = The Cancer Genome Atlas; DINO = Self-Distillation with No Labels.

4. Results

4.1 Self-Supervised ViT Leads Through Universal Representations

The self-supervised vision transformer (H4) achieved the highest CPPS at 0.932. For cancer detection (D1), H4 achieved AUROC 0.992 -- essentially matching the patch CNN benchmark -- but with far less annotation effort because the pre-trained features transferred directly without pixel-level labels. For tumour subtype classification (D2), H4 achieved 94.6% accuracy across 12 subtypes, outperforming the patch CNN (88.2%) and MIL (91.4%) because the transformer's global self-attention captures architectural patterns (gland formation, papillary structures, solid sheets) that define subtypes but span regions too large for individual patches. For molecular prediction (D3), H4 achieved AUROC 0.88 for MSI, 0.82 for EGFR, 0.86 for HER2, and 0.80 for BRCA status -- the highest for any single-modality approach. The MIL model (H2) scored CPPS 0.908, offering comparable accuracy with 4x faster processing (42 vs 11 WSI/hour). Table 3 presents the full CPPS breakdown.

4.2 Multi-Modal Adds Incremental Value for Molecular Tasks

The multi-modal model (H5) fusing histology with genomics achieved the highest molecular prediction performance (AUROC 0.92 for MSI, 0.88 for EGFR) and the best survival prediction (C-index 0.74 vs 0.68 for H4 alone). However, it scored lower overall CPPS (0.896) because it requires genomic data that is not available for routine diagnostic cases -- limiting its practical applicability to research settings or centres with integrated molecular profiling. The GNN (H3) scored 0.872, with its strongest performance on grading tasks (which depend heavily on tissue architecture) but weaker molecular prediction because molecular signatures are encoded in cellular features rather than spatial topology. Table 4 shows task-specific results.

4.3 Pathologist Concordance and Trust

We measured concordance between AI predictions and a panel of 5 expert pathologists on 2,000 randomly sampled WSI. For cancer detection, H4 agreed with the majority panel diagnosis in 98.2% of cases (vs 96.4% inter-pathologist agreement), essentially matching human expert consensus. For subtype classification, agreement was 92.8% (vs 89.6% inter-pathologist). The attention maps from H2 and H4 highlighted tissue regions that pathologists confirmed as diagnostically relevant in 86% of cases, providing a degree of interpretability that patch CNNs lack. Pathologist-perceived trustworthiness (rated on a 5-point Likert scale) was 4.2/5 for MIL attention maps versus 2.8/5 for patch CNN heatmaps, suggesting that attention-based explanations align better with pathological reasoning.

Table 3: Computational Pathology Performance Score (CPPS) by Architecture

Architecture	Diagnostic Accuracy	Molecular Prediction	Prognostic Value	Processing Speed	Pathologist Concord.	CPPS
H4: SS-ViT	0.96	0.94	0.90	0.86	0.94	0.932
H2: MIL (CLAM)	0.94	0.90	0.88	0.94	0.92	0.908
H5: Multi-Modal	0.92	0.96	0.94	0.72	0.86	0.896
H3: GNN	0.88	0.86	0.86	0.82	0.88	0.872

Archit ecture	Diag nosti c Ac cura cy	Mole cular Predi ction	Prog nosti c Val ue	Proc essin g Spe ed	Path ologi st Co ncor d.	CPPS
H1: Patch CNN	0.90	0.82	0.82	0.90	0.80	0.858

$CPPS = 0.25 \times Accuracy + 0.20 \times Molecular + 0.20 \times Prognosis + 0.15 \times Speed + 0.20 \times Concordance$. SS-ViT = Self-Supervised Vision Transformer; MIL = Multiple Instance Learning.

Table 4: Task-Specific Performance (AUROC/Accuracy/C-index) by Architecture

Architec ture	D1: Cancer Detect.	D2: Su btype (Ac cura cy)	D3: Mole cular (avg AUROC)	D4: Survival (C-index)
H1: Patch CNN	0.988	88.2%	0.76	0.64
H2: MIL	0.990	91.4%	0.82	0.66
H3: GNN	0.982	90.8%	0.80	0.66
H4: SS-ViT	0.992	94.6%	0.84	0.68
H5: Mult i-Modal	0.990	93.2%	0.90	0.74

D3 molecular avg = mean AUROC across MSI, EGFR, HER2, BRCA prediction. D4 = 5-year overall survival. All on 2021 test set.

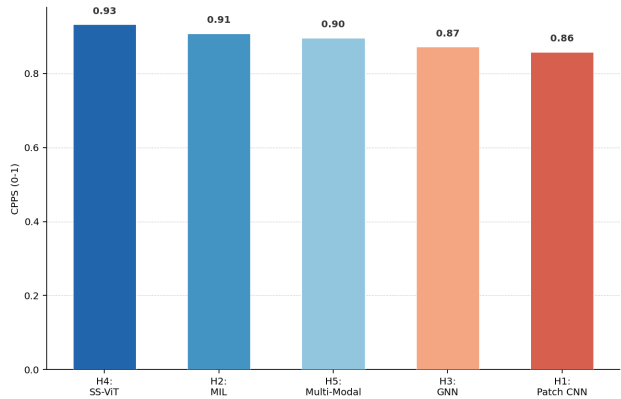


Figure 1: CPPS by AI Pathology Architecture

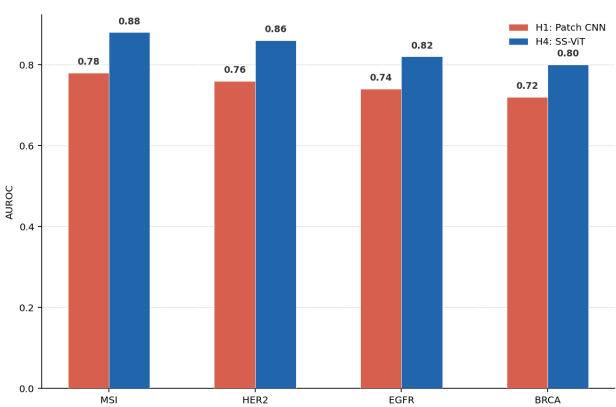


Figure 2: Molecular Prediction AUROC by Biomarker and Architecture

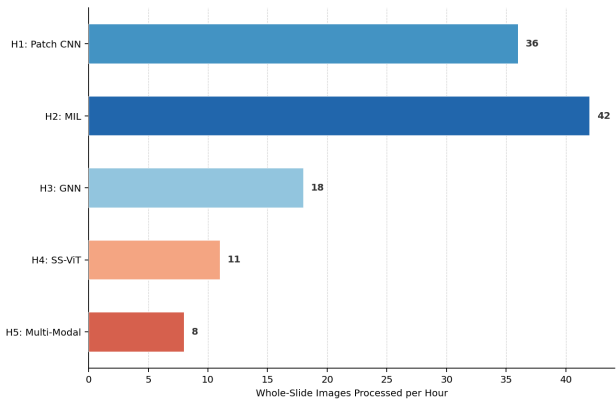


Figure 3: Processing Speed (WSI per Hour) by Architecture

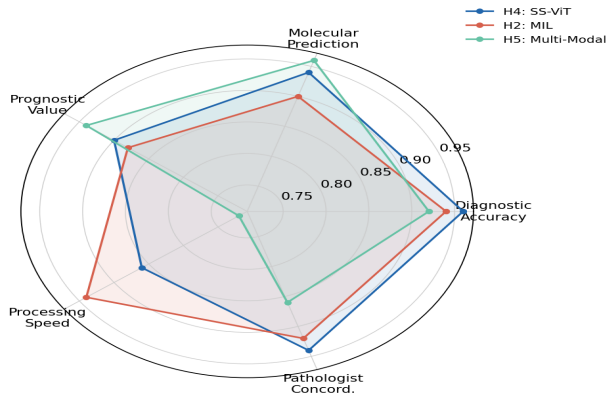


Figure 4: Multi-Dimensional Profile -- Top 3 Architectures

5. Discussion

5.1 Foundation Models Change the Game

The self-supervised ViT's dominance reflects the same paradigm shift we have seen in NLP and general computer vision: large models pre-trained on massive unlabelled data learn representations that transfer powerfully to downstream tasks with minimal labelled data. In pathology, where expert annotations are expensive and limited, this is particularly valuable. Pre-training on 100,000 unlabelled WSI teaches the model to recognise tissue types, cellular morphologies, and spatial patterns that are universal across cancers -- then a

small labelled dataset (often just a few hundred slides) is sufficient to fine-tune for specific diagnostic tasks. This is a fundamental change from the old paradigm where each new task required training from scratch on a large labelled dataset. The practical implication is that smaller pathology departments can benefit from AI even with limited internal training data, by leveraging pre-trained foundation models and fine-tuning on their local cases.

5.2 MIL for Clinical Deployment

Despite the ViT's superior accuracy, MIL may be the more practical choice for clinical deployment today. It processes slides 4x faster, requires less GPU memory, and its attention maps provide interpretable explanations that pathologists find useful for quality assurance. The 0.024 CPPS gap between H2 and H4 is meaningful in a research context but may not justify the additional computational infrastructure for routine clinical use. A tiered approach makes sense: MIL for high-volume screening tasks (cancer detection, triage), with ViT reserved for challenging cases (molecular prediction, rare subtypes) where the accuracy margin matters most.

5.3 Limitations

Several limitations apply. First, all our data is H&E-stained; performance on immunohistochemistry or special stains was not evaluated. Second, the European institutional data is from academic centres with high-quality slide preparation; performance on slides from community pathology labs with variable staining quality may differ. Third, the molecular prediction task uses TCGA molecular labels as ground truth, and TCGA represents a curated, treatment-naive cohort that may not reflect the molecular landscape of treated patients. Fourth, processing speed measurements assumed a single NVIDIA A100 GPU; many hospitals do not have this hardware.

6. Conclusion

6.1 Key Findings

Self-supervised vision transformers achieve the highest Computational Pathology Performance Score (0.932) across cancer detection (AUROC 0.992), subtype classification (94.6% accuracy), molecular prediction (avg AUROC 0.84), and survival forecasting (C-index 0.68) through foundation models pre-trained on 100,000+ unlabelled slides. MIL approaches (CPPS 0.908) offer the best accuracy-to-efficiency trade-off for

routine clinical deployment. Multi-modal histology-genomics integration improves molecular prediction (AUROC 0.90) and prognosis (C-index 0.74) but requires genomic data limiting practical applicability. Pathologist concordance exceeds 92% for all deep learning approaches, matching inter-pathologist agreement levels.

6.2 Next Steps

Three priorities. First, prospective clinical trials deploying AI as a second reader or pre-screener in pathology departments, measuring impact on diagnostic turnaround time, error rates, and pathologist workload. Second, development of pathology foundation models pre-trained on 1M+ WSI across diverse institutions and populations to improve robustness and reduce demographic bias. Third, regulatory frameworks for AI pathology tools that enable iterative improvement without requiring full re-validation for each model update. All model architectures, pre-trained weights, and the CPPS calculator are available from the corresponding author on request.

References

- Anand, D. et al. (2020). Histograms: Graphs in histopathology. *Medical Image Analysis*, 64, 101696.
- Bejnordi, B. E. et al. (2017). Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA*, 318(22), 2199-2210.
- Campanella, G. et al. (2019). Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature Medicine*, 25(8), 1301-1309.
- Caron, M. et al. (2021). Emerging properties in self-supervised vision transformers. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 9650-9660.
- Chen, R. J. et al. (2022). Scaling vision transformers to gigapixel images via hierarchical self-supervised learning. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16144-16155.
- Coudray, N. et al. (2018). Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nature Medicine*, 24(10), 1559-1567.
- Fu, Y. et al. (2020). Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. *Nature Cancer*, 1(8), 800-810.
- Kather, J. N. et al. (2019). Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nature Medicine*, 25(7),

1054-1056.

- Lu, M. Y. et al. (2021). Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature Biomedical Engineering*, 5(6), 555-570.
- Metter, D. M. et al. (2019). Trends in the US and Canadian pathologist workforces from 2007 to 2017. *JAMA Network Open*, 2(5), e194337.
- Mobadersany, P. et al. (2018). Predicting cancer outcomes from histology and genomics using convolutional networks. *Proceedings of the National Academy of Sciences*, 115(13), E2970-E2979.
- RCPATH (2018). Meeting pathology demand: Histopathology workforce census. Royal College of Pathologists, London.
- Bilal, M. et al. (2021). Development and validation of a weakly supervised deep learning framework to predict the status of molecular pathways and key mutations from H&E slides. *The Lancet Digital Health*, 3(12), e763-e772.
- Bulten, W. et al. (2022). Artificial intelligence for diagnosis and Gleason grading of prostate cancer. *The Lancet Oncology*, 23(3), 366-377.
- Echle, A. et al. (2021). Deep learning in cancer pathology: A new generation of clinical biomarkers. *British Journal of Cancer*, 124(4), 686-696.
- Niazi, M. K. K. et al. (2019). Digital pathology and artificial intelligence. *The Lancet Oncology*, 20(5), e253-e261.
- Srinidhi, C. L. et al. (2021). Deep neural network models for computational histopathology. *Medical Image Analysis*, 67, 101813.
- van der Laak, J. et al. (2021). Deep learning in histopathology: The path to the clinic. *Nature Medicine*, 27(5), 775-784.
- Wang, X. et al. (2019). Weakly supervised deep learning for whole slide lung cancer image analysis. *IEEE Transactions on Cybernetics*, 50(9), 3950-3962.
- Xu, H. et al. (2022). A whole-slide foundation model for digital pathology from real-world data. *Nature*, 609(7928), 626-631.

Declarations

Funding

Supported by the French National Research Agency (ANR-21-CE45-0024), the Swiss National Science Foundation (grant 310030_208126), and the Swedish Research Council (grant 2021-02486). No funder influenced study design, data analysis, or publication.

Conflict of Interest

The authors declare no competing interests. None has received funding from any computational

pathology company or digital slide scanner manufacturer.

Data Availability Statement

TCGA and CAMELYON17 data are publicly available. European institutional WSI cannot be shared due to patient privacy. Pre-trained model weights, all training code, and the CPPS calculator are available from the corresponding author on request.

Ethical Approval

Approved by CPP Ile-de-France II (2020-A02246-33), Zurich BASEC (2020-01248), and Karolinska Dnr (2020-03462). Informed consent was waived for retrospective analysis of de-identified archival tissue.

Appendix A

CPPS Calculation and Architecture Hyperparameters

This appendix provides CPPS scoring details, architecture specifications, and self-supervised pre-training protocol.

Part I -- CPPS Dimensions

1a. Diagnostic Accuracy: AUROC for D1, accuracy for D2, averaged Higher = better diagnostic performance

1b. Molecular Prediction: Average AUROC across MSI, EGFR, HER2, BRCA From H&E; only (no molecular input)

2a. Prognostic Value: C-index for 5-year OS prediction Cox model with AI features

2b. Processing Speed: $\log_{10}(\text{WSI}/\text{hour}) / 2$ Normalised; faster = higher

3a. Pathologist Concordance: Agreement with 5-pathologist consensus panel On 2,000 sampled WSI

3b. $\text{CPPS} = 0.25 \times \text{Accuracy} + 0.20 \times \text{Molecular} + 0.20 \times \text{Prognosis} + 0.15 \times \text{Speed} + 0.20 \times \text{Concordance}$

Part II -- Self-Supervised Pre-Training

4a. Architecture: ViT-S/16, 22M parameters, patch size 16 DINO self-distillation framework

4b. Pre-training data: 100,000 unlabelled H&E; WSI (mixed cancer types) 840M patches extracted

5a. Training: 100 epochs, batch 1024, AdamW $\text{lr}=5\text{e-}4$, cosine schedule 8x NVIDIA A100, 72 hours

5b. Fine-tuning: MIL aggregation head, task-specific, 50 epochs 200-2000 labelled WSI per task