ORGANIZED BY

ASDP
ASIAN SOCIETY OF DIGITAL PATHOLOGY

CO-ORGANIZED BY

IAP
International Academy of Pathology
Singapore Division

SingHealth DukeNUS
ACADEMIC MEDICAL CENTRE
PATHOLOGY

# ASDP 2026 SINGAPORE

3RD ANNUAL CONGRESS OF
ASIAN SOCIETY OF
DIGITAL PATHOLOGY

## INTELLIGENCE IN EVERY PIXEL:
## POWERING NEXT GENERATION PATHOLOGY

# ABSTRACT E-BOOKLET

ASDP 2026 SINGAPORE

# Session Abstracts

**Prof Jing Zhang**
Qiushi Distinguished Professor at
Zhejiang University
China

**Title**:

The Digital Brain: Shaping the Future of Precision Diagnosis in Neurological Diseases—Insights from Alzheimer's Disease

**Abstract** :

The rapid advancement of artificial intelligence (AI), digital pathology, and multi-omics technologies is transforming neurological disease research and diagnosis. Alzheimer's disease (AD), with its complex pathology and marked molecular heterogeneity, remains challenging to detect early, classify accurately, and stratify biologically. This presentation will examine how the emerging concept of the digital brain is reshaping precision neurodiagnostics by creating computational representations of brain pathology that integrate histomorphology, spatial molecular architecture, cellular composition, and fluid biomarkers across scales. Rather than treating pathology images, omics data, and clinical measurements as separate data streams, the digital brain framework seeks to reconstruct disease as a multidimensional, spatially organized biological system that can be quantitatively interrogated, compared, and ultimately linked to individual patients. Recent advances in AD research will be used to illustrate this approach, including AI-powered pathology foundation models capable of extracting clinically meaningful and potentially previously unrecognized features from complex tissue images, and the integration of spatial transcriptomics and other multi-omics technologies to define disease-associated cellular states, vulnerable brain regions, and molecular networks. Emerging approaches for detecting brain-derived extracellular vesicle biomarkers in peripheral blood will further demonstrate how information encoded in the brain may be projected into minimally invasive liquid biopsies for early diagnosis and longitudinal monitoring. By connecting computational models of the brain with neuropathology, molecular profiling, and peripheral biomarkers, the digital brain provides a framework for moving from descriptive diagnosis toward biologically informed disease classification, individualized risk assessment, and mechanism-guided clinical decision-making. This convergence of digital and molecular technologies may establish a new paradigm for precision diagnosis and personalized management of neurological diseases.

**Dr Lit-Hsin Loo, Ph.D**
Senior Principal Investigator
A*STAR Bioinformatics Institute (BII)
Co-founder, ImmunoQs
Singapore

**Title:**
From Slides to Spatial Intelligence: AI-Enabled Technologies for Spatial Profiling of Tumor Microenvironment

**Abstract** :

Cancer progressions and responses to therapy are shaped not only by cancer cells, but also by the spatial organizations of the entire tumor microenvironments. Immune cells, stromal cells, blood vessels, and tumor cells interact in highly structured neighborhoods. These architectures contain clinically important information that is difficult to capture by conventional histopathology assays. I will introduce computational technologies developed in our lab to transform multiplex tissue images into quantitative, spatially resolved maps of cancer. These include the HistoPath Analytics (HPA) platform for visualizing and scoring large histology and multiplex tissue images, CellShape AI for AI-driven cell segmentation in complex tumor tissues, and other automated approaches for scoring immune infiltration, tumor–stroma boundaries, and cellular neighborhoods across whole-slide images. I will highlight applications in gastric cancer, colorectal cancer, and nasopharyngeal carcinoma, where spatial profiling of the tumor microenvironment can reveal biological patterns linked to prognosis, disease progressions, and treatment responses. I will also discuss some on-going efforts, including same-slide spatial multi-modal profiling and foundation models for spatial proteomics, which will make tissue analysis more scalable, reproducible, and biologically interpretable. Together, these advances point toward a future in which AI-enabled spatial profiling helps uncover new cancer biology, guide biomarker discovery, and ultimately improve outcomes for patients.

**Dr Marilyn Bui**
Senior Member/ Professor
Departments of Pathology and Department of Machine Learning
Moffitt Cancer Center & Research Institute in Tampa, FL
United States of America

**Title:**
Augmenting Precision Pathology Through Digital Innovation and AI

**Abstract:**
Pathologists are the intellectual engine of precision oncology, delivering definitive diagnoses and actionable biomarker insights that guide life-saving therapies. Though often unseen by patients, they are the "doctor's doctor," shaping nearly every cancer care decision. Today, pathology stands at a pivotal inflection point. Rising diagnostic complexity, escalating case volumes, and a growing workforce shortage demand a new paradigm. Digital pathology and augmented intelligence (AI) are not optional upgrades. They are transformative tools redefining how pathology is practiced.

This presentation is designed for pathologists, trainees, machine learning scientists, and allied health and science professionals. It will explore the current landscape of digital pathology and AI adoption in the United States, highlighting real-world applications that enhance diagnostic accuracy, streamline workflows, and elevate reporting standards. It will also showcase practical resources developed by the digital pathology community to accelerate implementation.

My mission is to empower and inspire pathologists and their collaborators to lead the digital transformation and harnessing digital pathology and AI to advance precision medicine and deliver meaningful, measurable improvements in patient care.

**Patricia Diana Prasetiyo**
Faculty of Medicine
Pelita Harapan University; Siloam Hospitals Lippo Village
Indonesia

**Title**
Digital Pathology in Indonesia: Current Landscape and Future Directions

**Abstract**
Indonesia's archipelagic geography and uneven distribution of anatomical pathologists create challenges in access to diagnostic expertise. Digital pathology offers opportunities to support consultation, education, and diagnostic services across geographically dispersed healthcare settings. This presentation reviews the current landscape and future directions of digital pathology in Indonesia, drawing on public initiatives, emerging Indonesian research, and institutional experience.

The discussion considers selective digital use, hybrid diagnostic workflows, and the requirements for routine digital practice. Local experience with artificial intelligence (AI)-assisted HER2 assessment at Siloam Hospitals Lippo Village illustrates how image analysis can be incorporated into pathologist review. Tissue and staining quality, appropriate selection of invasive tumour, and critical assessment of AI output remain central to interpretation, with the pathologist retaining responsibility for the final report. Training activities and ongoing HER2 reporting consensus development are discussed alongside the need for representative local validation.

Implementation challenges include initial and recurring costs, infrastructure reliability, data storage and protection, laboratory information system integration, and workforce readiness. Sustainable adoption requires clear clinical governance, defined responsibilities, and evaluation of performance in the intended setting. Priorities include documenting active clinical implementations, establishing regional consultation networks, and conducting collaborative validation studies that reflect Indonesian laboratory conditions. Service evaluation should measure diagnostic quality, turnaround time, access to expertise, and cost per completed case. Collaboration among hospitals, universities, professional societies, and technology partners will be essential to translate digital capacity into reliable pathology services and broader access to specialist expertise.

**Keywords:** Digital pathology; artificial intelligence; HER2 assessment; telepathology; Indonesia.

**Prof Ruby Huang, Jun Yu**
Director, SMHI, International College
National Taiwan University
Professor
School of Medicine, College of Medicine
National Taiwan University
Taiwan

**Title:**
Mapping Tumor Heterogeneity in Ovarian Clear Cell Cancer

**Abstract:**
Ovarian cancer is a diverse disease group of histological subtypes with distinct clinical trajectories. Among these, ovarian clear cell carcinoma (OCCC) is showing high incidence in East Asia with critical unmet needs for novel therapeutic options. It is a histological subtype with distinct pathological features, molecular profiles, and biological functions. A major challenge for advances in OCCC is to translate the heterogeneity into actionable clinical decisions. To close this gap, we have established a multi-dimensional research framework for systemic mapping of tumor heterogeneity in OCCC. At the histo-pathological level, there exists profound intra-tumoral morphological heterogeneity. OCCC tumors frequently exhibit a complex admixture of papillary, tubule-cystic, and solid architectural patterns within a single slide. Our spatial proteomic profiling study has suggested that in early-stage OCCC with recurrence, there is an increased occurrence of tumor cells harboring immune-mimicry and selected morphological features surrounded by different lineages of tumor-infiltrating immune cells (TIIs). However, identifying this heterogeneity has not been straightforward in standard histopathological assessment, owing to inter-observer variability which makes consistent characterization of the architectural triad difficult to achieve at scale. Here, we developed ClearPath, a spatially resolved digital pathology pipeline that reframes the task from slide level aggregation to patch level classification on lock-down pathology foundation model features, enabling multi-class morphological classification and precise spatial localization directly within OCCC whole slide images. We conclude that a simple classifier on low magnification pathology foundation features could generalise across institutions without adaptation, and that the binding constraint on external performance is representation level domain shift together with case level diversity of the rare morphologies, not model capacity, class imbalance, or optimisation.

**Dr Yosep Chong, MD, PhD, MIAC**
Professor of Pathology
Eulji University College of Medicine
South Korea

**Title:**
LLM-Powered Pathology AI Agent: Automating Case Preparation in a Digital Pathology Workflow

**Abstract**
Pathology case preparation—extracting clinical histories, imaging findings, operative notes, and laboratory results into structured summaries and pre-filled synoptic reports—is a labor-intensive task traditionally delegated to senior pathology residents. We describe the design and early deployment of an LLM-powered pathology AI agent integrated within a real-world digital pathology workflow.

The AI agent operates as an ETL-to-LLM pipeline: patient data from HIS, LIS, and RIS are extracted via HL7/FHIR-compatible interfaces, normalized into structured JSON, and processed through a multi-step prompt-chaining pipeline. The agent generates (1) a clinical summary with present illness, clinical impression, and pertinent findings, and (2) a pre-filled synoptic reporting form conditionally modified based on specimen type, operative procedure, and attending pathologist preferences. A deterministic rule engine supplements the LLM to reliably trigger conditional fields (e.g., skin/nipple involvement for mastectomy, Miller-Payne grading for post-neoadjuvant chemotherapy). AI-populated fields are color-coded with hover-tooltip annotations explaining the source and reasoning, and conflicting data across sources are displayed side-by-side for pathologist adjudication.

Beginning September 2026, the agent is being deployed at our institution alongside a newly adopted digital pathology system. We present early findings from breast, gastrointestinal, and genitourinary cases, discussing integration challenges, the hybrid cloud/on-premise LLM deployment strategy for patient data security, and pathologist feedback on usability and accuracy. This work demonstrates a practical pathway for academic medical centers in Asia to adopt agentic AI within digital pathology workflows.

**Keywords:** large language model, AI agent, digital pathology, synoptic reporting, clinical workflow automation, prompt chaining

**Dr. Yuhao Yi**
Researcher
Institute of Clinical Pathology
West China Hospital, Sichuan University
China

**Title:**
CorePath: A Breast-Specialized Pathology Foundation Model for Breast Core Needle Biopsy Diagnosis and Risk-Controlled Report Generation

**Abstract**
Breast core needle biopsy (CNB) is central to breast cancer diagnosis but remains challenging because limited tissue sampling, lesion heterogeneity, and subtle morphologic overlap can obscure clinically important subtype distinctions. We developed CorePath, a breast-specialized multimodal pathology foundation model fine-tuned from PRISM using 7,903 paired CNB whole-slide images and diagnostic reports from two centers. CorePath was evaluated without task-specific retraining across six private multicenter CNB cohorts and two public breast pathology benchmarks. It outperformed the baseline models across cancer detection, invasion assessment, and histological subtyping in most private-center cohorts. On public benchmarks, CorePath achieved the highest weighted AUCs of 0.7780 for BCNB three-class invasive carcinoma subtyping, 0.8178 for BRACS three-class lesion stratification, and 0.8252 for BRACS seven-class fine-grained classification. In report generation, CorePath reduced non-breast hallucinations from 30.1% to 2.8%, indicating that breast-specific adaptation substantially improves domain fidelity. CorePath-CRG further combined conformal subtype-confidence gating with Learn-Then-Test risk control to enable selective report release, subtype-level fallback, and deferral. With this calibration strategy, CorePath-CRG produced no non-breast hallucinations in the evaluated cohorts and achieved the strongest overall performance on quantitative report-generation metrics and large language model-based report quality scores in most centers. These results support domain-specialized and statistically risk-controlled foundation models as a promising approach for accurate and reliable breast CNB diagnosis and report generation.

ASDP 2026 SINGAPORE

# Oral Presentations

## 36 - Contrast-Enhanced CT Radiomics with Pathology Foundation Model Distillation for Renal Cell Carcinoma Subtyping

**Yansong Zhang[1], Qingrong Sun[2], Maode Lai[1], Qiang Huang[3], Chong Lai[4]**

[1] The Department of Pathology and Pathophysiology of Zhejiang University School of Medicine
[2] Zhejiang Shuren University
[3] Department of Radiology, The First Affiliated Hospital, Zhejiang University School of Medicine
[4] Department of Urology, The First Affiliated Hospital, Zhejiang University School of Medicine

**Background/Objective:**
Renal cell carcinoma (RCC) encompasses three major histological subtypes — clear cell (KIRC), papillary (KIRP), and chromophobe (KICH) — each requiring distinct therapeutic strategies. While postoperative pathology remains the gold standard for subtyping, preoperative differentiation relies primarily on contrast-enhanced CT. We developed a four-phase CT radiomics model for RCC subtyping and explored whether knowledge distillation (KD) from a whole-slide image (WSI) foundation model could further improve CT-based classification.

**Methods:**
A total of 424 paired four-phase contrast-enhanced CT and WSI cases were analyzed (KICH 77, KIRC 267, KIRP 80). The CT model employed OmicsNet, a share-private encoder architecture operating on 1,502-dimensional radiomics features per phase with 14 HU/contrast augmentations. The WSI teacher model first extracted 768-dimensional patch features via TITAN, a vision-language pathology foundation model pretrained on 335,645 WSIs; patch features from pathologist-annotated tumor regions were aggregated through Gated Attention MIL to produce slide-level subtyping (AUC 0.9915). Soft labels from this teacher were then used to guide CT student training via KL divergence. All experiments followed identical 10-seed × 5-fold stratified cross-validation splits.

**Results:**
The CT radiomics baseline achieved an AUC of 0.9541 for three-class subtyping. KD consistently improved performance to AUC 0.9667, with positive gains across all 10 seeds. Improvements were most pronounced for seeds with lower baseline performance.

**Conclusion:**
Four-phase CT radiomics provides accurate preoperative RCC subtyping. KD from a WSI foundation model offers consistent improvement and enhanced robustness, demonstrating the potential of cross-modal knowledge transfer from pathology to imaging. The KD-enhanced model requires no pathology input at inference and remains clinically deployable.

# 47 - Impact of AI assistance on MASH fibrosis scoring in community pathologists

**Desiree Abdurrachim[1], Denise Ng[1], Charlene Ong[1], Serene Lek[2], Aileen Wee[3], Gwyneth Soon[3], Timothy J Kendall[4], Gideon Ho[2], Chih-Liang Chin[5], Saswata Talukdar[5], Richard Baumgartner[6], Samuel S Engel[7], Arun J Sanyal[8], David E Kleiner[9], Asad Abu Bakar Ali[1]**

[1] Quantitative Biosciences, MSD
[2] HistoIndex Pte. Ltd.
[3] Department of Pathology, National University Hospital
[4] Centre for Inflammation Research, Institute for Regeneration and Repair, University of Edinburgh
[5] Cardiometabolic Diseases, Merck & Co., Inc.
[6] Biostatistics and Research Decision Sciences, Merck & Co., Inc.
[7] Global Clinical Development, Merck & Co., Inc
[8] Stravitz-Sanyal Institute for Liver Disease and Metabolic Health, Virginia Commonwealth University School Of Medicine
[9] Laboratory of Pathology, National Cancer Institute, NIH

**Background:**
Inter- and intra-reader variability is a recognized challenge in liver biopsy scoring in metabolic dysfunction-associated steatohepatitis (MASH). While AI-assistance has been shown to improve inter-pathologist concordance among expert hepatopathologists, its utility in community pathologists with less hepatology experience remains unclear. Here, we evaluated the utility of AI-assistance in community pathologists (<5 years of hepatology experience) compared with expert hepatopathologists (15-30 years of experience).

**Methods:**
We used a Second Harmonic Generation (SHG)-based AI digital pathology platform (HistoIndex) that provided SHG images, AI-derived qFibrosis continuous scores and qFibrosis stages, in addition to standard H&E and Masson's Trichrome images. Pathologists read 120 liver biopsies (F0-F4; from NCT04583423 trial and Mount Sinai), with or without AI-assistance, separated by a 4-week washout period.

**Results:**
AI-assistance tended to improve inter-pathologist concordance in both the expert cohort (P=0.058) and the community cohort (P=0.096). The increase in kappa was numerically greater in the community cohort (0.18 ± 0.06) than in the expert cohort (0.13 ± 0.06). Scoring changes with AI-assistance were also larger in the community cohort (38%–60% of cases) than in the expert cohort (33%–42%). Correlation between pathologist scores and qFibrosis continuous values increased more substantially in the community cohort (Spearman's R: 0.77 without AI vs. 0.94 with AI) than in the expert cohort (0.80 without AI vs. 0.89 with AI). Using the median score of expert hepatopathologists as the gold standard, community pathologists showed a trend toward improved concordance with the gold standard (agreement P=0.09, kappa P=0.14), suggesting that AI-assistance may help narrow the gap between community and expert readers.

**Conclusion:**
AI-assistance shows promise for improving fibrosis scoring in MASH, particularly among community pathologists. However, the stronger impact on the community cohort compared to experts highlights the need for guidance to prevent over-reliance and ensure that pathologists continue to apply independent judgment.

# 56 - PathCoD: Pathology Chain-of-Discussion for Concept-Guided Multi-Agent Whole Slide Image Diagnosis

**Ashim Dhor, Rasel Mondal**
Indian Institutes Of Science Education And Research Bhopal

**Background:**
Whole slide image (WSI) classification in computational pathology remains challenging due to morphological heterogeneity, weak supervision, and limited interpretability of existing deep learning models. While pathology foundation models (FMs) provide strong representations, most existing approaches rely on monolithic architectures that insufficiently capture complementary diagnostic evidence across tissue patterns. We propose PathCoD, a concept-guided collaborative multi-agent reasoning framework for interpretable and robust WSI diagnosis.

**Methods:**
PathCoD integrates knowledge from multiple pathology FMs: WSI-LLaVA, CONCH, H-OptimusO, and Quilt-LLaVA. Teacher embeddings from these FMs are projected into a shared latent space and decomposed using non-negative matrix factorization to extract meaningful morphology concepts and validated using reconstruction error, silhouette score, and expert pathologist assessment. Specialized lightweight agents are then trained to focus on distinct pathology patterns such as nuclear morphology, tissue architecture, stromal microenvironment, inflammatory infiltrates, necrosis, and normal tissue. During inference, a query-aware routing mechanism dynamically selects the most relevant agents for a given case. Selected agents exchange latent evidence through the Chain-of-Discussion reasoning process, enabling collaborative refinement of predictions before confidence-weighted aggregation.

**Results:**
PathCoD was evaluated on TCGA (1,626 WSIs, 32 cancer subtypes) and externally validated on CAMELYON16 for metastasis detection. PathCoD achieved state-of-the-art performance on both datasets (TCGA: 89.2% Acc, 0.887 F1, 0.048 ECE; CAMELYON16: 94.7% Acc, 0.978 AUC) with lower computational complexity than recent pathology vision-language and multi-agent systems. Ablation studies showed concept-guided specialization, query-aware routing, and collaborative reasoning improved performance, while the discovered morphology concepts aligned with histopathology patterns verified by expert pathologists. Collaborative reasoning improved accuracy and calibration over static ensembling. Detailed results are provided in the supporting document.

**Conclusion:**
PathCoD combines concept-guided morphology specialization with collaborative multi-agent reasoning for interpretable and computationally efficient WSI diagnosis. PathCoD demonstrates strong generalization across diverse pathology tasks, suggesting collaborative AI as a promising direction for next-generation digital pathology and diagnostics.

# 59 - Deciphering the Immune Landscape of Parathyroid Carcinoma: Implications for Immunotherapeutic Responsiveness

**Md Ali Osama[1], Saransh Kanujia[1], Jayati Sarangi[1], Saumitra Choudhury[2], Riyaz Mir[3], Sonali Dixit[1], Millo Tabin[4], Shipra Agarwal[1]**

[1] Department of Pathology, All India Institute Of Medical Sciences
[2] Confocal Microscopy Facility, Centralized Core Research Facility, All India Institute of Medical Sciences
[3] Department of Biochemistry, All India Institute of Medical Sciences
[4] Department of Forensic Medicine & Toxicology, All India Institute of Medical Sciences

**Background:**
Parathyroid carcinoma is a rare endocrine malignancy with limited therapeutic options. Immune checkpoint blockade targeting the PD-1/PD-L1 axis has transformed the management of several solid tumors; however, its role in parathyroid carcinoma remains largely unexplored. Given that the response to immunotherapy is influenced not only by PD-L1 expression but also by the tumor immune microenvironment (TIME), characterization of the immune landscape in parathyroid carcinoma may provide insights into its potential therapeutic vulnerability.

**Objective**
To evaluate PD-L1 expression and comprehensively characterize the tumor immune microenvironment in parathyroid carcinoma, enabling identification of features predictive of immunotherapeutic responsiveness.

**Methods:**
Using immunohistochemistry and multiplex immunofluorescence, this ambispective study probed parathyroid carcinoma, parathyroid adenoma, and normal parathyroid tissues for PD-L1 expression (clone SP263, Ventana), infiltrating immune-cells, namely cytotoxic T lymphocytes (CTLs; $CD3^+CD8^+$), regulatory T cells (Tregs; $CD4^+CD25^+FOXP3^+$), and monocytic myeloid-derived suppressor cells (MDSCs; $CD11b^+CD14^+$). Immune cell densities were quantified per $mm^2$, and the TIME was categorized into four types based on PD-L1 expression and the presence of CTLs: Type I ($PD\text{-}L1^+$/CTL-rich, adaptive immune resistance), Type II ($PD\text{-}L1^-$/CTL-poor, immune ignorance), Type III ($PD\text{-}L1^+$/CTL-poor, intrinsic induction), and Type IV ($PD\text{-}L1^-$/CTL-rich, immune tolerance).

**Results:**
While 30% of adenomas showed patchy PD-L1 positivity (Tumor proportion score, TPS 3 to 6%), all carcinomas and normal parathyroid tissues were PD-L1-negative (TPS <1%). Among immune cells, CTLs and MDSCs were more abundant in the TIME of parathyroid carcinomas than in adenomas and normal tissues. Tregs were rare or absent. Parathyroid carcinomas demonstrated immune-tolerant (Type IV) or immunologically ignorant (Type II) microenvironmental profiles.

**Conclusions:**
Parathyroid carcinomas predominantly exhibit either an immune-ignorant ("cold") phenotype or a PD-L1-independent immune-tolerant microenvironment. The results predict a limited benefit from single-agent PD-1/PD-L1 blockade and the need for myeloid-targeting and immune-priming multimodal strategies.

# 69 - The Ethical, Regulatory, and Practical Landscape of Artificial Intelligence in Pathology: A Scoping Review with Secondary Confidence-Interval Re-analysis (2021–2026)

**Utso Guha Roy[1], Maher Sughayer[2,3]**
[1] Theism Diagnostics
[2] King Hussein Cancer Centre
[3] Asian Society of Digital Pathology

**Background:**
Computational pathology AI is entering clinical workflows faster than the governance infrastructure required to manage it. Prior reviews have catalogued ethical principles but have not interrogated the statistical robustness of empirical bias findings on which policy now rests, nor benchmarked pathology fairness reporting against adjacent modalities.

**Objectives:**
To map the global pathology AI ethics landscape (2021–2026); re-estimate statistical uncertainty around landmark bias claims; characterise reporting transparency; and translate findings into a deployable governance framework for pathology departments.

**Methods:**
PRISMA-ScR–compliant scoping review across PubMed/MEDLINE, Google Scholar, arXiv, and grey-literature retrieval from FDA, EU AI Office, MFDS, WHO, and OECD. Dual-reviewer screening ($\kappa = 0.84$) yielded 64 included records. Three pre-specified secondary analyses: (1) bootstrap simulation (10,000 resamples) with four-tier evidence classification of disparity confidence intervals for three Vaidya et al. (2024) pathology AI bias findings; (2) five-axis reporting transparency matrix across landmark pathology and radiology fairness studies; (3) comparative regulatory radar across four jurisdictions on five framework axes.

**Results:**
Only the lung-subtyping disparity (AUROC +10.94 pp; 95% CI 7.86–14.03) retained intervals excluding zero; breast-subtyping (95% CI −0.25 to +5.96) and IDH1-mutation gaps (95% CI −3.2 to +38.7) did not. Mitigation interventions reduced fairness gaps by 10.6–91.1%, but only 7 of 12 reported external validation. Pathology fairness papers consistently released code while withholding fold-level prediction outputs, limiting independent verification, a pattern absent in radiology comparators. The EU AI Act scored highest on the regulatory radar (24/25); WHO/OECD enforcement scored 1/5.

**Conclusions:**
Several widely cited pathology AI bias claims rest on weaker statistical footing than acknowledged. We propose a 3-tier translational framework (foundational, operational, sustaining) and a 10-point operational ethics checklist as an immediately deployable governance instrument for pathology departments, ahead of EU AI Act high-risk obligations in August 2027.

# 71 - AI-Powered Digital Pathology for Differential Diagnosis of IBD and Infective Colitis (Intestinal Tuberculosis)

**Anuradha Sekaran[1], Rajender Patel[1], Rupa Banerjee[1], Pranali Sonpatki[2], Shashank Gupta[2], Udayan Kachchhi[2], Nameeta Shah[2], Guduru Venkat Rao[1], Duvvur Nageshwar Reddy[1], Neha Rafeeq[1]**

[1] AIG Hospitals (a Unit Of Asian Institute Of Gastroenterology)

[2] Amaranth Medical Analytics

**Background and Objectives**

Differentiating inflammatory bowel disease (IBD) from infective colitis, particularly intestinal tuberculosis (ITB), remains a major diagnostic challenge in tuberculosis-endemic regions due to significant clinical and histologic overlap. Histopathologic interpretation is subjective and prone to interobserver variability. We evaluated whether automated quantitative analysis of crypt architecture and cellular density on H&E-stained colon biopsies using the Amaranth DIGEST AI pipeline could aid in classification of IBD, subtype differentiation, and distinction from ITB.

**Methods**

Digitized colon biopsy whole-slide images were analysed using AI models for tissue segmentation, crypt segmentation, nuclei detection, and mitosis detection. An 87-case discovery cohort (63 IBD, 24 non-IBD) and an independent 206-case validation cohort were used for IBD classification. For differential diagnosis, 13,586 biopsy fragments from 952 patient visits were analysed, of which 8,086 passed quality filtering. Extracted features included spatial-uniformity metrics, crypt morphology parameters, and cell-density features. Logistic regression and gradient boosting classifiers with cross-validation and patient-level aggregation were applied.

**Results**

Fifteen histologic features, including crypt density, lymphoid aggregate density, crypt integrity, and neutrophil density, significantly differentiated IBD from non-IBD colitis, achieving 95% accuracy in the discovery cohort and 91% accuracy in validation. Among 132 pathologist-equivocal cases, 91 were reclassified as IBD by AI. At the patient level, the model achieved an AUC of 0.817 for IBD versus normal and 0.875 for IBD versus ITB. The UC/CD/Infective classifier achieved AUCs of 0.909 for UC, 0.826 for CD, and 0.818 for infective colitis.

**Conclusion**

The AI platform demonstrated robust diagnostic performance for IBD and infective colitis classification and provides biologically interpretable histologic features. It represents a promising adjunctive decision-support tool for gastrointestinal pathologists, particularly in diagnostically challenging cases.

# 83 - Ki-67 Morphometry in Digital Pathology: From Prognostic Marker to Companion Diagnostic — A Comprehensive Review of Algorithms, Software, and Validation Strategies

**Hoa Pham Hoang Ngoc[1,2], Yosep Chong[3], Mohammed Rizwan Alam[3], Tsandmaa Byambadorj[3], Kyung Jin Seo[3], Kwangil Yim[3]**

[1] Hospital of University of Medicine and Pharmacy, Linh Dam
[2] University of Medicine and Pharmacy, Vietnam National University
[3] Department of Hospital Pathology, College of Medicine, The Catholic University of Korea

**Background / Objective**

The Ki-67 proliferation index is a core biomarker in breast cancer, neuroendocrine neoplasms (NENs), and soft-tissue sarcomas. Following the MonarchE trial and FDA approval of abemaciclib for HR+/HER2− breast cancer using a Ki-67 ≥20% cutoff, Ki-67 has evolved from a prognostic marker into a treatment-selection biomarker. However, manual scoring remains limited by substantial inter-pathologist variability, particularly within the clinically decisive 5–30% "Gray Zone." We reviewed current Ki-67 digital image analysis (DIA) platforms and evaluated emerging validation strategies for companion-diagnostic (CDx)-grade deployment.

**Methods**

Seventeen commercial and open-source Ki-67 DIA platforms were analyzed using regulatory filings, peer-reviewed literature, and conference data from 2010–2025. Empirical performance was primarily assessed using a head-to-head benchmark of 239 gastric-cancer tissue microarray cores (Alam et al., 2025) and updated Vision Transformer (ViT)-based results from Byambadorj et al. (2025).

**Results**

Annotation strategy influenced Ki-67 quantification more strongly than algorithm selection itself. Whole-TMA annotation consistently underestimated Ki-67 values relative to pathologist consensus because of stromal and inflammatory-cell contamination, whereas hands-free epithelial selection produced substantially improved concordance. Although overall inter-pathologist agreement was excellent (ICC 0.91–0.93), agreement declined markedly within the Gray Zone (ICC 0.43–0.66). ViT-based architectures with explicit tumor-cell classification outperformed first-generation convolutional neural networks, reducing deviation from manual consensus by up to 77% in challenging annotation settings. Pre-analytical variability from fixation, staining, and scanning contributed cumulative error approaching 35–45%, frequently exceeding between-algorithm variability.

**Conclusion**

Ki-67 DIA is becoming increasingly important for CDx-grade standardization in digital pathology. We propose a three-pillar framework consisting of disease-specific validation cohorts, an international public benchmark dataset, and human-in-the-loop adjudication of Gray Zone cases. Full automation remains premature for treatment-selection decisions, but AI is already indispensable for standardization and workload reduction when deployed within a controlled annotation protocol.

# 91 - A Foundation-Model Pipeline for Perineural Invasion Detection in Oral Squamous Cell Carcinoma: A Frugal Innovation Approach

**Shanthi Velusamy[1], Prajwal Dange[2], Sreeram M P[2], Sulakshana M S[1], Srinath B S[3]**

[1] Department of Pathology, Sri Shankara Cancer Hospital And Research Centre

[2] Department of Head and Neck Surgery, Sri Shankara Cancer Hospital And Research Centre

[3] Department of Surgical Oncology, Sri Shankara Cancer Hospital And Research Centre

**BACKGROUND**

Perineural invasion (PNI) is an adverse prognostic marker in oral squamous cell carcinoma, laborious to identify manually and prone to inter-observer variability. Existing artificial-intelligence solutions require large annotated datasets and GPU infrastructure, limiting applicability in low- and middle-income countries. We developed a frugal, deployable PNI detection pipeline using freely available tools and minimal annotations.

**METHODS**

A frozen pathology foundation model (Phikon-v2 ViT-L) extracted features from H&E field images at 20x and 10x magnifications. Two logistic-regression classifiers (nerve detection; PNI classification) were trained on the embeddings. Development proceeded through iterative retraining cycles driven by structured pathologist feedback and targeted hard-negative mining from precise failure coordinates. Macenko stain normalisation was applied at training and inference. The CPU-only pipeline retrains in under a minute and was deployed via Gradio with pathologist feedback. Across both magnifications, 2,104 training patches were extracted from 662 source slides.

**RESULTS**

Across 147 independent external images at 20x and 10x magnifications, the deployed pipeline achieved 88.4% PNI accuracy. At 20x on a previously unseen external cohort (n=23), accuracy was 92.86% (sensitivity 100%, specificity 85.71%); cross-validated patch-level AUCs were 0.999 for nerve detection and 0.979 for PNI. At 10x, the deployed v8 classifier reached 87.1% accuracy on 124 images across five batches, a 5.6 percentage-point improvement over the previous deployment; leave-one-image-out PNI AUC 0.972. Targeted hard-negative mining from exact failure regions consistently outperformed random sampling, adding 40–50 patches per cycle. Cross-magnification generalisation was preserved (100% nerve detection on 48 2x/5x/10x images, no retraining).

**CONCLUSION**

A clinically informative PNI detector was developed without GPU infrastructure, commercial software, or large annotated datasets. Targeted hard-negative mining with as few as 42 corrective patches repaired fundamental conceptual gaps. This frugal, human-in-the-loop approach demonstrates that accurate diagnostic AI can be developed in resource-constrained settings and adapted by LMIC institutions using locally annotated data.

## 92 - Artificial Intelligence-Assisted Diagnosis for Microinvasive Breast Carcinoma

**Yumian Jia[1], Jiatai Lin[2], Fangfang Liu[1], Yang Yang[1], Yunwei Han[1], Lili Zhai[1], Winmin Kong[1], Lei Guan[1], Jin Wang[1], Guangshen Gao[1], Liu Yang[1], Yuping Zhong[3], Jingxuan Zhou[3], Jingxing Luo[3], Kun Cao[3], Bingchao Zhao[2], Tianpeng Deng[3], Hannah Y. Wen[4], ZaiYi Liu[2], Chu Han[2], Xiaojing Guo[1]**

[1] Tianjin Medical University Cancer Institute & Hospital
[2] Guangdong Provincial People's Hospital
[3] South China University of Technology
[4] Memorial Sloan Kettering Cancer Center

**Background/Objective:**
Manual diagnosis of microinvasive breast carcinoma (MIBC) with extensive ductal carcinoma in situ (DCIS) is labor-intensive and subjective. This study aims to develop and validate a deep learning-based Artificial Intelligence （AI） model to automatically identify and quantify microinvasive foci, thereby augmenting pathologists' diagnostic efficiency and enabling novel quantitative insights.

**Methods:**
We developed MicroDet, a deep learning system combining epithelial segmentation, myoepithelial cell detection, and microinvasion classification. A novel tissue re-staining protocol generated precisely aligned H&E/IHC (p63/Cytokeratin) whole-slide image pairs for training. The training set included 166 WSI pairs from 99 patients (48 DCIS-MIBC, 51 invasive carcinoma). Validation used 928 WSIs from 82 DCIS-MIBC cases.

**Results:**
MicroDet achieved ROI-level AUCs of 0.932 (training) and 0.917 (validation). In a reader study with six breast pathologists (junior, intermediate, senior) reviewing 30 challenging cases (187 WSIs), AI assistance significantly streamlined diagnosis. Mean diagnostic time decreased by 142 seconds per case (patient-level) and 22 seconds per WSI (slide-level) across all pathologists, with consistent gains across experience levels. Additionally, automated quantification revealed that the number of AI-detected microinvasive foci was significantly higher in cases with lymph node metastasis compared to those without metastasis ($p < 0.001$), suggesting a potential quantitative histomorphometric biomarker.

**Conclusion:**
MicroDet accurately detects MIBC, improves pathologist efficiency, and provides an objective measure of microinvasion burden that correlates with lymph node metastasis. This study highlights AI's potential for both workflow optimization and discovery of novel quantitative biomarkers in surgical pathology.

# 101 - Spatial Validation of Tumor–Stroma Ratio in Oral Squamous Cell Carcinoma: From Visual Hotspot Assessment to Whole-Slide Digital Quantification

**Mimna V. M[1], Priyanka Srivastava[1], Ruchika Aggarwal[2]**
[1] Department of Pathology, All India Institute Of Medical Sciences (AIIMS), Gorakhpur, Up, India
[2] Department of ENT, All India Institute of Medical Sciences (AIIMS), Gorakhpur

**Background/Objective:**
Tumor Stroma Ratio (TSR) serves as a prognostic marker in oral squamous cell carcinoma (OSCC) and is conventionally assessed visually at a single, pathologist-selected hotspot. However, the extent to which this region represents the tumor stroma remains unclear. This study aims to validate automated TSR quantification using hematoxylin and eosin (H&E) and pan cytokeratin (panCK) staining against manual annotation, assess spatial representativeness across tumor regions, and compare interobserver agreement with automated assessment.

**Methods:**
Forty resected OSCC specimens were analyzed. The most stroma-rich tumor section from each specimen was selected for assessment and panCK staining. H&E and panCK whole slide images were analyzed using QuPath. TSR was assessed by visual estimation, manual annotation, and a trained pixel classifier, with panCK manual annotations as a reference standard. Visual assessment was independently performed by both an experienced and a trainee pathologist. Manual and automated TSR quantification was conducted in the highest TSR region (hotspot), the invasive front, and the whole tumor region, and results were correlated with tumor-infiltrating lymphocytes, tumor budding, depth of invasion, and pathological stage. Statistical analyses were conducted using R software.

**Results:**
Visual assessment of TSR showed moderate interobserver agreement, while automated quantification exhibited stronger agreement with the experienced pathologist (ICC > 0.85). PanCK staining delineated the tumor-stroma interface more clearly than H&E, especially in keratinized and inflamed regions. Hotspot TSR showed moderate concordance with both invasive-front and whole-tumor TSR. Among the evaluated regions, invasive-front TSR demonstrated the strongest correlation with adverse histopathological parameters, including tumor budding and depth of invasion.

**Conclusion:**
Automated digital quantification of the TSR offers an objective and reproducible approach compared to manual assessment. Spatial heterogeneity in TSR was observed across tumor regions, with evaluation at the invasive front demonstrating significant prognostic value in OSCC.

# 112 - An AI-Based Whole Slide Image Analysis Pipeline for Screening of Head and Neck Mucosal Biopsies: Development and Independent Validation at a High-Volume Cancer Centre

**Swapnil Rane[1], Shubham Makwana[5], Tripti Bameta[1,3], Anshuman Sharma[1], Saurabh Nagar[1,2], Mrudula Katarni[1,2], Shireen Ali[1,2], Jaykumar Patel[5], Meera Santosh[5,6]**

[1] Dept. of Digital & Computational Oncology, Tata Memorial Centre-ACTREC
[2] Dept. of Pathology, Tata Memorial Centre-ACTREC
[3] Homi Bhabha National Institute
[4] Dept. of Pathology, Tata Memorial Centre-TMH
[5] The Innovation Story - HealthCare ( TIS-H)
[6] Gleneagles Hospital

**BACKGROUND / OBJECTIVE**

Head-Neck squamous cell carcinoma (HNSCC) accounts for a disproportionate burden of cancer mortality in South Asia, where biopsy volumes are high, and pathologist capacity is constrained. AI-assisted whole slide image (WSI) analysis offers a scalable approach to triage, yet most published models lack validation on real-world diagnostic case-mix. We developed and validated a multi-model, multi-output AI pipeline for screening H&N mucosal biopsies on H&E-stained WSIs.

**METHODS:**

The pipeline comprises a tissue detection module, a CLAM-based semi-supervised tumor detection model, an artifact-detection model, a parallel cut detection model, an attention-to-segmentation conversion module, and a pixel-level tissue-type detection model. The core CLAM-based model was trained on WSIs from 650 patients undergoing biopsy for suspected HN cancer, with near-equal distribution of malignant(SCC) and non-malignant diagnoses, using 10-fold cross-validation. The pipeline generates the following outputs: three-class (invasive cancer, suspicious/atypical proliferation, non-neoplastic) tissue- and WSI-level classifications, along with a probability score; a patch-level heatmap visualization; and pixel-level tumor segmentation. Independent validation was performed on an additional 1,863 patients, reflecting real-world diagnostic distribution, including hyperplasia, leukoplakia, granulation tissue, and inflammatory lesions.

**RESULTS:**

In the 10-fold cross-validation & independent cohorts respectively, the pipeline achieved high test AUC(0.9508,0.9619), sensitivity(91.8%,93.6%), specificity(86.9%,87.3%), PPV(96.07%,96.2%), and NPV(74.46%,79.8%) for the invasive & non-neoplastic categories. Heat maps enable explainability and review of relevant areas within the suspicious/atypical proliferation categories. Most false positives were flagged using the blur-detection module. Common causes of false negatives were the absence of subepithelial connective tissue (flagged using the tissue-type detection module) and an extremely well-differentiated tumor.

**CONCLUSION:**

This multi-model pipeline demonstrates robust, explainable, clinically reliable accuracy for HNSCC screening on a large, real-world validation cohort. Its multi-output design and multi-model workflow reduce FPR & FNR, supporting integration into routine pathology workflows. Prospective multi-institutional clinical validation in consecutive biopsies, integration into the clinical WSI viewer, and impact assessment are ongoing.

# 117 - Systematic Evaluation of Diagnostic Accuracy, Errors, and Hallucinations in Large Language Models for Diagnostic Pathology

**Kris Lami[1], Aleksandra Asaturova[2], Serdar Balci[3], Andrey Bychkov[4], Daniel Gomes Pinto[5], Agnes Harahap[6], Jennifer Kim[7], Thiyaphat Laohawetwanit[8], Santosh Menon[9], Jijgee Munkhdelger[4], Hoa Pham[10], Ayushi Sahay[11], Swati Satturwar[12], Kurumi Seki[4], Maher Sughayer[13], Yuri Tachibana[4], Ilknur Turkmen[3], Qingqing Wu[12], Junya Fukuoka[1,4]**

[1] Nagasaki University Graduate School of Biomedical Sciences
[2] FSBI «National Medical Research Center for Obstetrics, Gynecology and Perinatology named after Academician V.I.Kulakov» Ministry of Health of the Russian Federation Pirogov Russian National Research Medical University
[3] Memorial Healthcare Group, Pathology Laboratory
[4] Kameda Medical Center
[5] Pathology Department, Hospital Garcia de Orta
[6] Department of Anatomical Pathology, Faculty of Medicine, Universitas Indonesia/Dr. Cipto Mangunkusumo Hospital
[7] Department of Tissue Pathology & Diagnostic Oncology, ICPMR, Westmead Hospital, New South Wales Health Pathology
[8] Division of Pathology, Chulabhorn International College of Medicine, Thammasat University
[9] Department of Pathology, Urology and Gyanecology Disease Management Group, Homi Bhabha National Institute, Tata Memorial Centre
[10] Department of Medical Laboratory, Hospital of University of Medicine and Pharmacy
[11] Department of Pathology, Tata Memorial Hospital (TMH), Tata Memorial Centre
[12] Department of Pathology, Wexner Medical Center, The Ohio Sate University
[13] Department of Pathology and Laboratory Medicine, King Hussein Cancer Center

**Background:**
Large language models (LLMs) are increasingly explored in diagnostic pathology, but most evaluations emphasize diagnostic accuracy alone. How these models fail, particularly through pathology-relevant errors and hallucinated histologic features, remains underexplored, despite their importance for clinical safety.

**Objective:**
To evaluate diagnostic accuracy, overall performance, and error/hallucination patterns of four LLMs across multi-organ pathology cases.

**Methods:**
Four LLMs (GPT-5.2, Gemini 3, Grok 4.1, and Opus 4.6) were evaluated using 118 retrospective diagnostic pathology cases from two medical institutions, covering 15 organs. No data split, model training, or fine-tuning was performed. For each case, LLMs generated microscopic descriptions and diagnoses from multi-magnification microscopic images with clinical context. Eighteen pathologists evaluated outputs using a structured scoring sheet. Diagnostic correctness, error type, hallucination type, and overall performance on a 1–5 scale were assessed. Overall scores were compared using the Friedman test with Wilcoxon signed-rank post hoc tests.

**Results:**
A total of 472 LLM outputs were analyzed, with 118 outputs per model. Overall strict diagnostic accuracy was 52.1%, and partial-or-better accuracy was 62.3%. Strict accuracy was 51.7%, 55.9%, 39.8%, and 61.0% for LLM1–4, respectively, corresponding to the models listed above. Mean overall scores were 3.44 ± 1.29, 3.44 ± 1.46, 2.89 ± 1.49, and 3.48 ± 1.43, respectively. Overall scores differed significantly among models ($p < 0.001$), with LLM3 performing significantly worse than LLM1, LLM2, and LLM4. Any error was present in 79.9% of outputs and any hallucination in 72.9%. The most frequent error was misinterpretation (70.1%), followed by omission (54.2%). The most frequent hallucination was fabricated feature (71.4%), followed by unsupported inference (20.4%).

**Conclusion:**
LLMs showed moderate diagnostic accuracy but frequent pathology-relevant errors and hallucinations. Misinterpretation and fabricated histologic features were dominant failure modes, highlighting the need for systematic error-based validation before clinical use in pathology.

# 120 - DeepLNM: A Multimodal Model for Predicting Lymph Node Metastasis in Colorectal Cancer

**Zhengyu Zhang[1,2], Chengyu Lu[1,2], Zhijian Cen[1,2], Li Liang[1,2,3]**

[1] Department of Pathology, Nanfang Hospital, Southern Medical University

[2] Guangdong Province Key Laboratory of Molecular Tumor Pathology

[3] Jinfeng Laboratory

**Objective:**

Accurate assessment of lymph node metastasis (LNM) is essential for prognostic stratification and treatment selection in colorectal cancer (CRC). However, results based on imaging and routine pathology remain vulnerable to false-negative and false-positive findings. Here, we developed DeepLNM, a multicenter multimodal foundation model for LNM prediction in CRC, and rigorously evaluated it using retrospective, external, prospective, and user-study evidence to support clinical decision-making in real-world pathology workflows.

**Methods:**

DeepLNM was trained and evaluated using a multimodal dataset collected from three hospitals across northern and southern China. The dataset includes matched H&E whole-slide images (WSIs), CT scans, and text, with retrospective data further supplemented by WSIs and diagnostic information from TCGA. Model performance was assessed across internal and external cohorts, followed by prospective validation in a registered clinical trial. To interpret model behavior, we performed visualization and dimensionality-reduction clustering to reveal LNM-associated pathological patterns. Generalization was additionally tested on an independent multimodal gastric cancer cohort.

**Results:**

DeepLNM achieved a mean AUC of 0.922 in the internal cohort and AUCs of 0.872 and 0.848 in 2 external validation cohorts; in the prospective cohort, DeepLNM achieved an AUC of 0.845. In a user study, integrating DeepLNM improved diagnostic accu- racy, particularly among junior pathologists (94.1% to 97.5%), and reduced diagnostic time while increasing diagnostic confidence. In an independent gastric cancer cohort, DeepLNM generalized to LNM prediction (AUC, 0.793). Visualization and clustering analyses identified histopathologic features associated with LNM.

**Conclusion:**

DeepLNM is a validated multimodal decision-support tool that improves LNM assessment across diverse evaluation settings, including clinical evidence and user performance. The model's ability to generalize beyond CRC and its interpretability via feature visualization support its potential integration into pathology practice to enhance diagnostic accuracy, efficiency, and ultimately patient care.

# 122 - More Experts, Less Accuracy? Optimizing Multi-Expert Ground Truth Construction for Clinically Valid AI in Pathology

**Makoto Kawamoto[1], Frederico Gaia[1], Ethan Okoshi[1], Shintaro Sato[2], Kris Lami[1], Junya Fukuoka[1]**
[1] Department Of Pathology Informatics, Nagasaki University
[2] Saitama Red Cross Hospital

**Background**
Clinical application of artificial intelligence (AI) requires high accuracy and reproducibility. Because AI models reproduce predefined Ground Truth (GT), GT reliability directly affects clinical validity. In pathology, where substantial interobserver variability exists, optimal GT construction remains unclear. This study aimed to identify effective strategies for multi-expert GT construction by comparing consensus methods across four independent cohorts.

**Methods**
Four datasets with independent expert annotations and clinically relevant outcomes were analyzed. Cohorts A and B involved UIP classification in interstitial pneumonia (9 and 28 experts, respectively), Cohort C involved lung adenocarcinoma subtype classification (18 experts), and Cohort D involved granuloma subtyping (13 experts). Consensus GTs were generated using: (1) majority vote of all raters; (2) majority votes of randomly selected expert subsets (n = 3, 5, 7); and (3) hierarchical clustering of raters followed by cluster-specific majority voting (G1), or Delphi refinement when clustering was not observed (G2). GTs were evaluated against survival (A, C), disease progression (B), or diagnostic discrimination (D).

**Results**
Across Cohorts A–C, GTs generated from all experts generally showed weaker associations with clinical outcomes than alternative approaches. In random-subset analyses, smaller expert groups demonstrated stronger outcome correlations (mean HR = 4.99 for n = 3 vs 3.25 for n = 7), suggesting that increasing the number of experts may dilute clinically relevant signals. Cluster-based consensus consistently achieved the strongest correlations across cohorts. In Cohort D, Delphi refinement improved agreement ($\kappa = 0.27$ to 0.47) but yielded only modest gains in diagnostic discrimination ($p = 0.001$ to $6.55 \times 10^{-6}$).

**Conclusions**
Increasing the number of expert raters does not necessarily improve clinical validity. Consensus strategies based on expert clustering produced the most clinically relevant GTs, while Delphi refinement provided limited benefit for non-clusterable categories. These findings support strategic GT construction in pathology and fields with high interobserver variability.

# 127 - Morphology-Guided Multimodal Deep Learning for Improving Diagnostic Concordance of Invasive Lobular Carcinoma: A Multicenter Study

**Chang Liu[1], Qi Liu[2], Jiaxian Miao[1], Si Wu[1], Zhongze Cui[1], Yueping Liu[1]**
[1] The Fourth Hospital Of Hebei Medical University
[2] School of Information Engineering, Hebei GEO University

**Background:**
Distinguishing invasive lobular carcinoma (ILC) from invasive ductal carcinoma (IDC) of the breast and accurate histological subtyping of ILC remain challenging in routine clinical pathology practice, accompanied by unsatisfactory interobserver diagnostic agreement, highlighting an urgent need for standardized artificial intelligence-assisted diagnostic tools. Methods: We developed a pathology-text guided clinical multimodal fusion model termed Multimodal mOrphology-guided Reasoning with PHenotype Optimization (MORPHO). By incorporating morphological prior knowledge and multimodal clinical data, the proposed model enables robust differentiation between ILC and IDC as well as interpretable subclassification across four ILC histological subtypes. A multicenter retrospective cohort was established, consisting of an internal dataset from the HCH center (1,409 patients with 2,262 whole-slide images [WSIs]) and six independent external validation cohorts (262 patients with 351 WSIs in total). MORPHO adopts a dual-branch network architecture: the image branch leverages the MedSigLIP encoder pre-trained on large-scale pathology image-text pairs to adaptively select the top 8% most morphologically relevant regions from each WSI guided by diagnostic textual descriptions, followed by feature aggregation and classification via the DTFD-MIL framework; the clinical branch applies XGBoost to model structured clinical metadata. A logistic regression-based decision-level fusion strategy integrates predictive outputs from two branches to yield final patient-level diagnoses. Results: In internal test cohorts, MORPHO achieved outstanding performance for binary ILC/IDC discrimination (AUROC=0.985, accuracy=0.953) and favorable efficacy for four-category ILC subtyping (AUROC=0.889, accuracy=0.796). Consistently robust generalization performance was validated across all six external independent datasets. Notably, a reader performance study involving nine pathologists with varying seniority demonstrated prominent concordance improvement aided by MORPHO: the Fleiss' Kappa coefficient for ILC/IDC differential diagnosis increased from 0.514 to 0.859, while the Kappa value for ILC subtyping rose from 0.504 to 0.763. Collectively, our findings verify that MORPHO recapitulates pathologists' morphological diagnostic reasoning and substantially alleviates inter-individual diagnostic variability.

# 141 - A Clinically Validated Foundation Model for Comprehensive Lung Pathology Interpretation

**Zhengrui Guo[1], Zhengyu Zhang[2,3], Li Liang[2,3], Hao Chen[1]**
[1] Hong Kong University of Science and Technology
[2] Department of Pathology, Nanfang Hospital, Southern Medical University
[3] Department of Pathology, School of Basic Medical Sciences, Southern Medical University

**Background/Objective**
Pathological assessment is fundamental to lung cancer care, yet current computational pathology approaches rely on task-specific models for isolated objectives, and pan-cancer foundation models lack subspecialty-level depth and have not been comprehensively evaluated across clinical workflows or prospectively validated. We introduce PulmoFoundation, a multi-center, prospectively validated, randomized controlled trial (RCT)-evaluated foundation model for comprehensive lung pathology assessment across pre-operative, intra-operative, and post-operative care.

**Methods**
PulmoFoundation was built upon Virchow2 via lung-specific continual self-supervised pretraining using 88 million tiles from 39,503 H&E-stained whole slide images (WSIs) across 12 sources. It was evaluated on 26,643 WSIs across 32 clinically relevant tasks spanning diagnostic assessment, biomarker prediction, and prognostication, with validation across 32 internal and 21 external cohorts from 8 independent institutions. We further conducted a registered prospective observational study of 1,357 consecutive patients across 11 diagnostic tasks in routine practice and a crossover RCT with 8 pathologists evaluating 616 cases under AI-assisted and unassisted conditions.

**Results**
PulmoFoundation achieved clinical-grade performance in core diagnostic tasks across biopsy, frozen-section, and surgical-resection slides, outperforming four pan-cancer foundation models in retrospective benchmarking. In the prospective cohort, PulmoFoundation achieved an average AUC of 92.3%. Using pre-specified high-confidence triage thresholds, PulmoFoundation could reduce second-review burden for 68.8% of biopsy and 83.0% of frozen-section cases, and defer 44.5% of IHC stain orders, with PPVs of 1.000, 0.991, and 0.966, respectively. In the crossover RCT, AI assistance improved diagnostic accuracy across 4,928 matched case-reader comparisons (91.7% vs. 83.8%; OR=2.23, 95% CI: 2.05-2.42; $p<0.001$), reduced median diagnostic time by 19.6%, increased diagnostic confidence by 8.7%, and improved inter-rater agreement from moderate (kappa=0.56) to substantial (kappa=0.76).

**Conclusion**
Together, these evaluations support PulmoFoundation as a clinically validated decision-support system for pulmonary pathology and show how subspecialty-specific foundation models can extend computational pathology from isolated endpoints toward integrated clinical care.

# 150 - Democratizing and accelerating AI-driven pathology research through agentic intelligence

**Jiabo MA, Cheng Jin, Yihui Wang, Hao Chen**
The Hong Kong University Of Science And Technology

**Background:**
Recent advances in computational pathology and foundation models have greatly expanded the scope of AI-driven biomedical discovery. However, translating clinical hypotheses into executable computational workflows remains technically challenging, requiring expertise in programming, machine learning, and pathology. We developed PathLab, an agentic AI framework that democratizes and accelerates computational pathology research by converting natural-language research objectives into validated end-to-end analytical workflows.

**Methods:**
PathLab employs a multi-agent architecture that translates user intent into executable computational pathology pipelines through structured orchestration of reusable domain-specific skills and tools. The framework incorporates pathology-specific guardrails to prevent methodological errors, including data leakage, task–data mismatches, and unsupported workflow configurations. We evaluated PathLab across 12 public datasets spanning four representative task families: region-of-interest classification, whole-slide image classification, segmentation, and survival prediction. Performance was compared with expert-designed pipelines using a pre-specified non-inferiority margin of −5%. Robustness, safety, usability, and workflow efficiency were further assessed through systematic stress testing and controlled user studies involving pathologists and AI researchers.

**Results:**
PathLab achieved non-inferior performance compared with expert-developed pipelines across all task categories, with a mean relative performance difference of −0.8% (non-inferiority $p < 0.001$). In safety evaluations, the framework successfully intercepted 99% of adversarial or methodologically invalid requests while maintaining a 99.0% zero-hallucination rate during workflow generation. User studies demonstrated substantial reductions in workflow construction time, enabling executable computational pathology studies to be generated within minutes through natural-language interaction. Furthermore, synergistic human–AI collaboration using PathLab achieved performance comparable to or exceeding expert-only implementations on several complex pathology tasks.

**Conclusion:**
PathLab provides a reliable interface between biomedical research intent and computational execution, enabling computational pathology studies to be designed at the level of scientific questions rather than programming expertise. By maintaining methodological rigor while dramatically reducing technical barriers, PathLab offers a scalable foundation for democratizing and accelerating AI-driven pathology research.

# 162 - TuroChat: An Agentic AI Copilot for Pathology

**Swapnil Bhat[1], Rajiv Kaushal[2], Jay Mehta[3], Prasenjit Das[6], Sheena Alphones[4], Apurv Shah[5], Maulik Shah[1], Devansh Lalwani[1]**

[1] Turocrates.ai
[2] Tata Memorial Hospital
[3] NM Medical
[4] Neuberg Oncopath
[5] Unipath Labs
[6] AIIMS

**Background:**
Synoptic reporting per College of American Pathologists (CAP) protocols improves data completeness, yet adoption remains inconsistent due to the time burden of template navigation and context-switching between quantitative tools. Agentic AI copilots that autonomously dispatch case-appropriate algorithms and draft structured reports may reduce this friction while preserving diagnostic authority.

**Methods:**
We conducted a two-centre, prospective, within-subject crossover pilot. Four board-certified pathologists each reported 20 de-identified cases in two arms: (A) conventional reporting and (B) TuroChat-assisted reporting with voice dictation. The 40-case library comprised 20 breast (10 resections, 10 core biopsies) and 20 GI specimens (8 colorectal carcinoma resections, 6 polyps, 6 gastric biopsies). TuroChat autonomously routed each case to the appropriate algorithm: Ki-67 quantification, ER/PR Allred scoring, and mitotic-figure detection (breast); MSIpredict for microsatellite instability triage (colorectal carcinoma); polyp classification with dysplasia grading; and H. pylori detection (gastric biopsies), then drafted a CAP-compliant synoptic report. The primary endpoint was median case-level turnaround time (TAT). Secondary endpoints included synoptic completeness, MSIpredict concordance, word error rate (WER), System Usability Scale (SUS), inter-pathologist agreement, and error rate on blinded re-review.

**Results:**
Results: Median case-level TAT was reduced by 40% with TuroChat-assisted reporting relative to conventional reporting ($p < 0.05$). Synoptic completeness was 30% higher with TuroChat assistance. WER was 5.3%. The orchestration layer dispatched the correct algorithm in 100% of cases (40/40). Mean SUS score was 90/100. No increase in diagnostic errors was identified on blinded re-review.

**Conclusion:**
TuroChat-assisted reporting reduced median TAT by 40% across breast and GI domains while maintaining diagnostic accuracy. A single orchestration layer routing diverse case types to specialised algorithms supports the feasibility of an agentic copilot architecture for AI-augmented pathology reporting.

# 166 - NucQuant: AI-powered user interface for nuclear stain quantification in immunohistochemistry

**Satyajit Ghosh[1], Sherin Mathew[1], Adity Solanki[2], Umesh Sherkhane[1], Anshuman Sharma[1], Surabh Nagar[1], Mrudula Katarni[1], Asawari Patil[1], Mohammed Shahin[1], Sagar Tripathy[1], Subhash Yadav[1], Tripti Bameta[1], Swapnil Rane[1]**

[1] ACTREC, Tata Memorial Centre, Kharghar

[2] Mumbai University

Objective. Nuclear immunohistochemistry (IHC) biomarker analysis is routinely used for cancer diagnosis and prognosis, but is still largely based on manual microscopic assessment, leading to limited reproducibility and inter-observer variability. AI-based automated analysis offers a promising approach to improve the efficiency, consistency, and reliability of biomarker quantification. Methods. This study used image patches from 118 IHC slides stained for Ki-67, ER, and PR with available expert diagnoses. An open-source deep learning model for nuclear IHC segmentation was fine-tuned for cell segmentation. Positive and negative cells were classified via HSV thresholding, followed by intensity-based scoring using optical density measurements with expert-defined thresholds and a fuzzy classification scheme. Allred score was computed and validated against clinical diagnoses. An end-to-end pipeline was implemented as a standalone Streamlit application, with both whole-image and region-of-interest analysis, along with a companion Python library reproducing core functionalities. Results. Transfer learning substantially improved segmentation performance on the test set, increasing mean Dice and IoU by 1.8× and 2.2×, respectively, compared with the vanilla model. The intensity classification module achieved F1 scores of 0.84, 0.65, and 0.87 for weak, moderate, and strong cells. Software-derived Ki-67 proportional scores demonstrated excellent agreement with two independent pathologist assessments (ICC(2,1) = 0.89 and 0.78), while the agreement between the two pathologists was ICC(2,1) = 0.77. Software-derived Allred scores showed moderate agreement with pathologist assessments (QWK = 0.59 and 0.52), compared with an inter-pathologist agreement of QWK = 0.69. The complete pipeline was deployed as a Streamlit application and is under prospective large-scale validation in our hospital. Conclusion. The proposed framework provides a reproducible and automated solution for quantitative analysis of nuclear IHC biomarkers, demonstrating good agreement with expert pathological assessment. To facilitate transparency, reproducibility, and further development by the research community, the source code will be publicly available on GitHub.

# 169 - Tracking Evolving Tumour Knowledge in Low-Resource Settings Using a Tagging System with Local LLM Support that Respects WHO Tumour Taxonomy

**Ruseik Rahumath[1], Pasindu Nanayakkara[2], Dilani Lokuhetty[3], Iciar Indave[3]**
[1] Ministry Of Health And Mass Media
[2] Postgraduate Institute of Medicine, University of Colombo
[3] International Agency for Research on Cancer (IARC), World Health Organization

**Background/Objective**
Tumour pathology knowledge evolves faster than systematic reviews can synthesise. In low- and middle-income countries (LMICs), workload constraints render continuous literature surveillance impractical. Large language models (LLMs) enable knowledge comparison but pose hallucination and cloud-dependency risks. We developed a low-cost, locally hosted, offline LLM-driven knowledge management platform for tumour literature surveillance, aligned to the WHO Classification of Tumours (5th Edition) taxonomy, designed for resource-limited pathology settings.

**Methods**
A web-based platform was developed by a clinician-led team at negligible cost using AI-assisted coding. Knowledge entities are curated and tagged via a multi-dimensional taxonomy aligned to the WHO Classification of Tumours, 5th Edition. A self-hosted LLM (Phi-4 Mini, 3.8B parameters) operates fully offline. Usability was evaluated using the System Usability Scale (SUS) administered to 50 consecutive usage instances by senior level postgraduate histopathology trainees. Comparison analyses are grounded exclusively on provided texts to mitigate hallucination, producing analytical narratives for expert review, preserving human interpretive control.

**Results**
Core components were successfully implemented: WHO-aligned hierarchical tagging, knowledge entity lifecycle management, visual Boolean search, LLM comparison workflow, and role-based access control. The system operates on 6–8 GB RAM without GPU dependency or internet connectivity. A pilot usability evaluation with senior level postgraduate histopathology trainees yielded a mean System Usability Scale (SUS) score of 75.6 (SD = 12.6, 95% CI [72.0, 79.2]), significantly exceeding the industry benchmark of 68 ($t(49) = 4.25$, $p < 0.001$), corresponding to Grade B (62nd percentile) on the Sauro–Lewis curved grading scale. Internal consistency was good (Cronbach's $\alpha = 0.866$).

**Conclusion**
A low-cost, offline LLM-driven platform for tumour pathology literature surveillance was developed, deployed, and validated with end-users in a resource-constrained LMIC setting. Integrating a locally hosted LLM with WHO-aligned tagging demonstrates a replicable model for responsible AI adoption in pathology.

# 171 - Strengthening Diagnostic Review and Quality Assurance Through a Low-Cost Telepathology Network in Resource-Limited Settings

**Prima Rachmita[1], Anandia Putriyuni[1], Santy Saberko[1], Dandy Citra[2], Ryan Haroen[3]**
[1] Ropanasuri Surgical Specialty Hospital
[2] Sitanala General Hospital
[3] Nusantara Capital City General Hospital

**Background:**
The use of telepathology in routine diagnostic practice has expanded globally, improving diagnostic efficiency and supporting patient safety. Consultation among pathologists plays an important role in ensuring diagnostic accuracy, particularly in challenging cases. However, expert consultation may not always be immediately accessible. This study aimed to evaluate the role of a low-cost telepathology consultation network in supporting diagnostic review and quality assurance.

**Methods:**
A retrospective study was conducted on pathology consultation cases discussed through a low-cost telepathology network between January and December 2024. Microscopic images and relevant clinical information were shared through widely available communication platforms among pathologists, including collaborations among pathologists located in different Indonesian islands. Consultation outcomes were categorised as diagnostic concordance, diagnostic revision, or recommendation for additional ancillary investigations.

**Results:**
A total of 168 consultation cases were reviewed. Cases represented a broad spectrum of pathology, including breast pathology (15.48%), cytology (13.10%), female reproductive (12.50%), soft tissue and bone (10.12%) accounting for the largest proportions. Diagnostic concordance with the initial diagnosis was observed in 77.38% of cases, while diagnostic revision occurred in 12.50%. Additional ancillary investigations were recommended in 10.12% of cases to establish a definitive diagnosis. All consultations were completed remotely without physical slide transfer or dedicated digital pathology systems. The network facilitated access to peer expertise across geographically separated regions and supported discussion of diagnostically challenging cases.

**Conclusion:**
A low-cost telepathology consultation network can effectively support diagnostic practice and quality assurance in resource-limited settings. The network reinforced diagnostic review, identified discrepant interpretations, guided ancillary investigations, and enabled access to peer expertise across geographically dispersed regions without physical slide referral. This model represents a practical and scalable approach to strengthening pathology services in areas where advanced digital pathology infrastructure remains inaccessible.

# 205 - FISH-Net: An Open Source Multi-Filter Approach for Automating Cytogenetics

**Praveen Sharma[1], Satyender Dharamdasani[1], Saaransh Sharma[2], Niti Shyamsukha[1], Vipul Sudhir Patil[1], Om Singh[2], Sukrit Gupta[4]**

[1] Department of Hematology, Postgraduate Institute Of Medical Education And Research
[2] Department of Computer Science Engineering (CSE), Indian Institute of Technology
[3] Department of Electronics and Communication Engineering (ECE), Punjab Engineering College
[4] Department of Biomedical Engineering, Indian Institute of Technology

**Background/Objective:**
Fluorescence in situ hybridization (FISH) is a key cytogenetic technique for detecting clinically significant chromosomal abnormalities in pathological disorders. Manual FISH analysis is repetitive, time-consuming, and subject to inter-observer variability. Standard RGB-based deep learning methods struggle with spectral overlap, particularly for fusion signal detection, which is critical for translocation diagnosis. Fusion signals help identify translocation mutations that drive oncogenesis, guide targeted therapies, and inform prognosis. We propose FISH-Net, a multi-filter signal detection pipeline designed to automate FISH interpretation with clinically relevant accuracy.

**Methods:**
FISH-Net analyzes individual DAPI, FITC, ORANGE, and AQUA channels separately using instance segmentation, followed by confidence-modulated geometric fusion detection to identify translocation signals. The pipeline was validated on a high-throughput clinical benchmark comprising approximately 4,500 nuclei from 368 patients, representing a diverse dataset of hematological cases. Performance was evaluated against a standard RGB-based baseline using signal-level detection metrics (F1, mAP50) and a count-based metric matching nucleus-level enumeration standards used in clinical translocation adjudication.

**Results:**
FISH-Net substantially outperformed the RGB-based baseline, achieving an overall F1 score of 0.91 compared to 0.21, with mAP50 of 0.93 across individual signal channels. Using the count-based metric reflecting nucleus-level enumeration for translocation adjudication, the method achieved an F1 of 0.74. A pilot deployment as a clinician-ready web application was implemented, providing nucleus-level reporting suitable for seamless integration into pathology workflows.

**Conclusion:**
FISH-Net is the first in-house instance-segmentation–based AI framework for automated FISH interpretation. By addressing spectral overlap through individual channel analysis and confidence-modulated fusion detection, it achieves substantial improvements over conventional RGB approaches. By reducing manual workload, shortening turnaround time, and improving reproducibility, FISH-Net has the potential to enhance patient care in high-volume and resource-limited diagnostic laboratories, while also contributing a unique open-source dataset for further research.

# 214 - Human-AI Collaboration Improves Agreement in Histologic Subtyping While Preserving Prognostic Relevance: a study in lung adenocarcinoma

**Frederico Gaia Costa da Silva, Kris Lami, Junya Fukuoka**
Nagasaki Univeristy

**Background:**
Interobserver variability remains a major challenge in lung adenocarcinoma subtyping and grading, particularly for complex histologic patterns such as micropapillary and acinar growth. It affects reproducibility of histological assessment and downstream prognostic stratification. We evaluated whether AI-assisted interpretation impacts diagnostic consistency as well as the prognostic relevance.

**Methods:**
This retrospective study included 98 lung adenocarcinoma resection cases (2009-2013). 12 pathologists (1 beginner, 8 junior, 3 senior) assessed histological subtype proportions in two sequential phases: human-only and AI-augmented interpretation, using WSIs with AI-generated inference maps from a previous validated independent deep-learning model, previously validated. No wash-out period applied, intentionally simulating real-world second-read decision support. Interobserver agreement was assessed using intraclass correlation coefficient (ICC) and Fleiss' kappa, and prognostic performance was evaluated using Kaplan-Meier analysis.

**Results:**
AI-assisted interpretation was associated with increased agreement in subtype quantification, particularly in less reproducible patterns. ICC improved most in micropapillary (0.385→0.610, 95%CI 0.295–0.495 to 0.519–0.705, p=0.002) and acinar patterns (0.428→0.598, 95%CI 0.334–0.539 to 0.500–0.697, p=0.024), with additional gains in lepidic and mucinous components. Agreement for predominant subtype increased (Fleiss' kappa: 0.42±0.07 → 0.60±0.07, z=3.57, p<0.001), and WHO grading agreement also improved (0.50±0.08 → 0.65±0.08, z=2.69, p=0.009). Survival stratification according to WHO grade was preserved across all expertise levels at both timepoints (log-rank p<0.05). Discriminatory performance assessed by Harrell's concordance index improved among beginners (0.678→0.711), remained stable among juniors (0.676→0.673), and remained high among seniors (0.734→0.710).

**Conclusion:**
AI-assisted interpretation reduced interobserver variability in lung adenocarcinoma subtyping and grading without compromising prognostic discrimination. By improving reproducibility while preserving the biological and clinical significance of established histologic classifications, AI may provide a valuable decision-support framework for routine digital pathology practice.

ASDP 2026 SINGAPORE

# Poster Presentations

# 11 - Automated Image Quality Benchmarking of Whole Slide Imaging Platforms for Digital Cytopathology: A Dual-Method Quantitative Approach Using Deep Learning and Classical Image Processing

**Yosep Chong, Tsandmaa Byambadorj, Seunghyun Choi**
College of Medicine, The Catholic University of Korea, South Korea

**BACKGROUND**
WSI adoption in cytopathology is hindered by the three-dimensional nature of cytologic specimens, which demands multiplane focusing and generates substantial data volumes. Conventional pathologist-based scanner validation introduces inter-observer variability and lacks scalability. Objective, automated quality control frameworks for cytology-specific WSI benchmarking remain lacking.

**METHODS**
Twelve cytopathology cases (eight organ sites; conventional smear and liquid-based cytology) were scanned at 40× under four conditions: single-layer and three-layer acquisition on 3DHistech FLASH and Vieworks LH210 (Realtime EF). Image quality was assessed by two complementary automated pipelines: (1) GrandQC, an EfficientNet-B0 UNet++ deep learning model generating seven-class artifact segmentation maps at 7× magnification; and (2) a Laplacian variance pipeline for focus quantification using Otsu-thresholded tissue segmentation and 512×512 pixel tiling. Artifact area proportions were compared with Wilcoxon signed-rank tests and Bland-Altman analysis; inter-method correlation was assessed by Pearson and Spearman coefficients.

**RESULTS**
LH210 showed significantly lower total artifact burden than 3DHistech under matched single-layer conditions (18.09% vs. 29.68%; $p = 0.005$), with markedly lower out-of-focus rates (0.42% vs. 2.16%). Bland-Altman analysis confirmed a systematic bias of 11.6 percentage points favoring LH210 (95% LoA: −9.0 to +32.2%). GrandQC detected 3–5× more artifact than the Laplacian pipeline overall, yet both converged on the same platform ranking. Inter-method correlation was low ($r = -0.24$; $\rho = -0.27$), confirming orthogonal quality dimensions. Critically, GrandQC misclassified 3D cytology cell clusters as pen markings—a domain shift from histopathology training data—highlighting the need for cytology-specific model adaptation. LH210 also achieved one-third of 3DHistech's scanning time and file size.

**CONCLUSIONS**
Dual deployment of deep learning and classical image processing provides a scalable framework for objective WSI benchmarking in digital cytopathology. Both methods confirmed LH210's artifact and efficiency advantages, while revealing a critical gap: histopathology-trained AI models require cytology-specific adaptation before reliable deployment.

## 13 - Cloud-Based Telepathology as a Global Equalizer: A Win–Win Model for Resource Optimization and AI Advancement

**Kirtikumar Pilankar**
MEDICHECK DIAGNOSTIC CENTER

**Background / Objective:**
Significant global disparities exist in pathology services, with resource-limited countries such as India facing high patient volumes and a shortage of trained pathologists, leading to delayed diagnoses. In contrast, developed nations possess advanced digital pathology infrastructure and artificial intelligence (AI) capabilities but often have lower case volumes per pathologist. This study explores cloud-based telepathology as a model to bridge this gap and create a mutually beneficial global framework.

**Methods:**
A conceptual and narrative analysis of global pathology workforce distribution, digital pathology adoption, and AI development was undertaken. Existing frameworks in telepathology, digital health strategies, and AI applications were reviewed to evaluate the feasibility of cross-border diagnostic collaboration using cloud-based platforms (1,2).

**Results:**
Cloud-based telepathology enables real-time sharing of histopathological images, facilitating expert consultation across geographical boundaries. This model improves diagnostic turnaround time and accuracy in high-burden regions while optimizing specialist utilization globally. Additionally, it generates access to large, diverse datasets from resource-limited settings, enhancing the development, validation, and generalizability of AI algorithms compared to homogeneous datasets from developed regions (3).

**Conclusion:**
Cloud-based telepathology represents a sustainable, equitable, and scalable solution for addressing global disparities in pathology services. By fostering ethical collaboration, data sharing, and capacity building, this model offers a win–win strategy—improving healthcare delivery in underserved regions while accelerating AI innovation globally. It has the potential to become a cornerstone of a more inclusive and efficient global healthcare ecosystem.

References 1.World Health Organization. Global strategy on digital health 2020–2025. Geneva: WHO; 2021. 2.College of American Pathologists. Validating whole slide imaging for diagnostic purposes in pathology. Arch Pathol Lab Med. 2013;137(12):1710–1722. 3.Geoffrey Hinton, et al. Deep learning. Nature. 2015;521:436–444.

## 15 - Launching Digital Pathology for Clinical Use in Jordan

**Maher Sughayer**
King Hussein Cancer Center (Jordan)

**Background:**
a previously conducted and published survey regarding the use of digital pathology in Jordan revealed the near absence of digital pathology use in standard diagnostic work up. It was used rarely for consultations only. One year ago, the King Hussein cancer Center decided to implement digital pathology for all purposes including use in routine case sign-out.

**Methods:**
after studying the options for equipment and infrastructure that will match our objectives an acquisition of the appropriate equipment and preparation of infrastructure were made. Installation validation of the system was made. Then validation of the system for diagnostic use was established based on the College of American Pathologists recommendations (CAP). Sixty cases for H&E and additional 20 cases for IHC were used for each pathologist after a washout period of at least 2 weeks. After the validation gradual implementation started with breast biopsies first, followed by GI biopsies and prostate biopsies. AI for breast biomarkers was introduced 9 months later.

**Results:**
initial hesitation from some pathologists was encountered, however using the gradual implementation has led to wide acceptance and the process is going smoothly with around 90% implementation in the mentioned fields. Ongoing implementation will continue until we achieve 100% implementation. Technical challenges are occasionally faced including system failure and poor slide quality involving around 10% of the cases leading to repeat scans.

**Conclusion:**
gradual implementation of digital pathology for clinical diagnosis use appears to be successful despite some challenges. This approach is highly recommended for the successful implementation of digital pathology. Other uses of digital pathology for education, consultation and multidisciplinary meetings has also improved.

# 17 - Artificial Intelligence-Enabled Continuous HER2 Assessment in the Antibody-Drug Conjugate Era

**Min Hsiang Chang[1], Hsin-Hsiu Tsai[2], Chun-Jui Chien[2], Jian-Chiao Wang[2]**
[1] Taipei Li Jen Pathology Clinic
[2] AI Lab, Quanta Computer lnc.

**Background:**
Antibody-drug conjugates (ADCs) redefined HER2 from an oncogenic driver to a drug delivery address, where even low-level expression enables targeted therapy. Categorical scoring now determines ADC eligibility, but pathologists show poor agreement at the very boundaries that matter most: 0 vs 1+ and 1+ vs 2+. Multiple groups have converged on AI-enabled continuous quantification as the necessary response.

**Method:**
Transformer-based AI classifies epithelial patches from standard HER2 IHC whole slide images into 0/1+/2+/3+ categories. C-score is derived as: C-score = $0 \times P_0 + 1 \times P_{1+} + 2 \times P_{2+} + 3 \times P_{3+}$, where $P_i$ = proportion of patches in each category (range 0–3), maintaining direct visual correspondence with standard assessment.

**Result:**
C-scores were evaluated in a balanced pilot cohort (n=66) across all HER2 categories. Unambiguous cases (0 and 3+) showed tight, homogeneous distributions, while borderline categories (1+ and 2+) displayed broad, overlapping ranges — reflecting a biological continuum poorly captured by discrete boundaries. Three data-driven thresholds were identified for clinically relevant binary tasks: T3=2.54 for anti-HER2 eligibility (AUC 1.00), T2=0.42 for ISH triage (AUC 0.96), and T1=0.14 for HER2-low identification (AUC 0.98).

**Conclusion:**
Unlike other continuous approaches (qIHC, QCS), c-score preserves direct visual-cognitive correspondence so pathologists can verify AI output using the same 0/1+/2+/3+ vocabulary they already use. The continuous measurement enables flexible threshold optimization for different clinical contexts: high-sensitivity screening, confirmatory testing, or emerging ADC indications. Beyond global scoring, the patch-based architecture inherently captures spatial heterogeneity, a dimension potentially critical for predicting ADC bystander effect. If validated against treatment outcomes, AI-driven continuous HER2 scoring transitions from optional enhancement to essential diagnostic infrastructure, paralleling how molecular testing became indispensable once EGFR-TKI pairing proved its clinical necessity. And we will walk with AI toward a future where diagnostic precision matches the evolving therapies.

# 18 - Comparative Evaluation of a Foundation Model and a Lightweight Transformer for Automated Her2 Scoring in Breast Cancer Whole Slide Images

**Min Hsiang Chang[1], Yeh-Han Wang[2], Hsin-Hsiu Tsai[3], Chun-Jui Chien[3], Jian-Chiao Wang[3]**

[1] Taipei Li Jen Pathology Clinic

[2] Department of Anatomical Pathology, Fu Jen Catholic University Hospital

[3] AI Lab, Quanta Computer lnc.

**Background:**

HER2 scoring is critical for modern breast cancer therapies, especially with emerging indications of antibody-drug conjugates for HER2-low tumors. However, interobserver agreement remains limited in borderline cases. AI-based scoring has the potential to improve diagnostic consistency and scalability. Materials and Methods: We developed and compared two transformer-based models for HER2 scoring on breast cancer whole-slide images (WSIs): a large-scale foundation model (Virchow) and a lightweight model (TinyViT). Both were trained using patch-level annotations and integrated into a WSI scoring pipeline. Performance was evaluated on an independent test set (n=66), including clinical decision tasks and inference efficiency.

**Results:**

Both models achieved substantial agreement with pathologist reports (linear weighted kappa: 0.860 for Virchow, 0.825 for TinyViT). Virchow showed slightly higher WSI-level accuracy, while TinyViT reduced inference time by 60%. In three binary clinical tasks, both models performed comparably to pathologists, particularly in identifying HER2-low tumors. A continuous scoringframework demonstrated a strong correlation between models (Pearson r = 0.995) and aligned with human assessments.

**Conclusion:**

Both transformer-based AI models achieved human-level accuracy in interpretable HER2 scoring. While the foundation model offers marginally higher accuracy, the lightweight model provides practical advantages for clinical deployment. In addition, continuous scoring may support more granular HER2 quantification, especially in borderline cases.

# 19 - AI-ASSISTED TISSUE COMPLETENESS AND TUMOR DETECTION IN MOHS SURGERY: AN ASIAN COHORT STUDY

**Wen-Bin Luo[1], Li Yang Loo[2], Isaac Lam[1,4], Seow Ye Heng[1], Fan Foon Cheo[1,3], John Huang[4], Wei Song Shang[4], Terrence Yeoh[4], Li Yan Khor[1,3], Chee Leong Cheng[1,3], Choon Chiat Oh[2,3], Po Yin Tang[1,3]**

[1] Department of Anatomical Pathology, Singapore General Hospital
[2] Department of Dermatology, Singapore General Hospital
[3] Duke-NUS Medical School
[4] AI Singapore

**Background/Objective:**
Mohs micrographic surgery (MMS) is the gold standard for skin cancers treatment, particularly for basal cell carcinoma (BCC), due to its precise real time margin assessment. However, histologic interpretation of multiple stages of frozen sections is labor-intensive and subject to interpretive variability. This study aims to develop and validate AI models for automated BCC detection and tissue section completeness assessment in MMS frozen sections in an Asian population.

**Methods:**
This retrospective single-center study was conducted at Singapore General Hospital. Archived MMS frozen section slides from 130 patients (majority Chinese) with BCC who underwent MMS (2017-2025) were digitized into whole slide images (WSIs). Tumor regions were annotated, and tissue sections were classified as complete or incomplete based on predefined histologic criteria. Convolutional neural network (CNN) models with attention mechanisms were trained for (1) patch-level BCC detection and (2) section-level tissue section completeness assessment. Performance was evaluated using sensitivity, specificity, and F1-score.

**Results:**
The study collected 734 WSIs in total. For BCC detection, the ConvNeXt model achieved an F1-score of 0.90, sensitivity of 0.91, and precision of 0.89 at the patch level. Our model was able to exclude BCC confounders including hair follicles, sweat ducts, and sebaceous glands. For tissue section completeness assessment, the ConvNeXt model achieved an F1-score of 0.56, sensitivity of 0.52, and precision of 0.61 at the section level. Our model was able to pick up incomplete epidermis integrity and incomplete deep margins representations.

**Conclusion:**
To our knowledge, this study is the first in an Asian population, demonstrating the feasibility of deep learning for simultaneous tumour detection and tissue completeness assessment in MMS frozen sections. Integration of AI tools can expedite BCC detection and support histologic margin evaluation, which warrants further investigation for its potential to improve surgical workflow timing, healthcare costs, as well as patients' treatment outcomes.

## 20 - AutoFNAC: A Closed-Loop AI-Guided Fully Automated Fine Needle Aspiration System Integrated with Real-Time Digital Cytopathology

**Kirtikumar Pilankar**
MEDICHECK DIAGNOSTIC CENTER- RATNAGIRI

Title: AutoFNAC: A Closed-Loop AI-Guided Fully Automated Fine Needle Aspiration System Integrated with Real-Time Digital Cytopathology Author: Dr. Kirtikumar Pilankar

**Background:**
Fine needle aspiration cytology (FNAC) is a rapid and minimally invasive diagnostic technique widely used for evaluating superficial and deep-seated lesions. However, it remains highly operator-dependent, with variability in sampling adequacy and diagnostic yield. Recent advances in artificial intelligence (AI), robotics, and Digital Pathology provide an opportunity to reimagine FNAC as a standardized, automated, and intelligent diagnostic workflow.

**Objective:**
To design a prototype for a fully automated, AI-integrated FNAC system capable of real-time lesion targeting, robotic sampling, automated cytology processing, and instant adequacy assessment.

**Methods:**
The proposed AutoFNAC system incorporates multimodal imaging using Ultrasound imaging and CT scan for lesion localization. AI-based segmentation models identify optimal sampling zones while avoiding necrotic regions. A robotic needle guidance unit performs precise aspiration with controlled depth and pressure. Extracted material undergoes automated liquid-based cytology preparation, staining, and whole-slide digitization. A deep learning cytology engine analyzes cellular morphology, provides a benign–malignant probability score, and evaluates specimen adequacy. A novel closed-loop feedback mechanism enables automatic repeat sampling from alternate sites if inadequacy is detected. Results (Expected): The system is anticipated to improve sampling adequacy rates, reduce operator variability, and enhance diagnostic accuracy compared to conventional FNAC. Additionally, it may decrease procedure time and minimize the need for repeat aspirations.

**Conclusion:**
AutoFNAC represents a next-generation paradigm in Cytopathology by integrating AI, robotics, and digital cytology into a unified, closed-loop system. This prototype has the potential to transform FNAC into a standardized, reproducible, and intelligent diagnostic modality, paving the way for future innovations such as real-time precision cytopathology and 4D diagnostic analytics.

# 23 - SMM-MIL: Spatially-aware Morpho-Molecular Multiple Instance Learning for Survival Prediction on Whole Slide Image

**Junhua Yu[1], Tianpeng Deng[2], Fangfang Liu[3], Jingqi Huang[4], Xiaojing Guo[3], Zhenwei Shi[5], Changhong liang[5], Zaiyi Liu[5], Wenbin Liu[1], Peng Xu[1], Chu Han[5]**

[1] Guangzhou University
[2] South China University of Technology
[3] Tianjin Medical University Cancer Institute and Hospital
[4] Southern Medical University
[5] Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences)

**Background**

Accurate survival prediction from whole slide images (WSIs) is critical for personalized patient management. Due to the scarcity of annotations on WSIs, multiple instance learning (MIL) has emerged as the effective weakly supervised approach for survival prediction. Although existing methods enhance prognostic performance by exploring spatial context among patches within the MIL framework, accurate prognosis analysis is highly dependent on molecular information. However, the lack of paired morpho-molecular data significantly limits comprehensive prognostic analysis. To address this challenge, we propose SMM-MIL, a spatially-aware morpho-molecular framework for survival prediction.

**Methods**

By incorporating an image-transcriptomic foundation model, our MIL framework can leverage its powerful representation capability to extract molecular-prior features directly from pathological patches, thereby overcoming the limitation caused by the absence of paired data. Furthermore, we design a series of modules to effectively discover morpho-molecular information while substantially improving overall training effectiveness and efficiency. First, we deploy the Disagreement-Guided Sampling (DGS) strategy to select patches exhibiting high cross-feature complementation potential, effectively mitigating the patch redundancy. Then, we utlize Molecular-Prior Gated Fusion (MGF) module to dynamically calibrate primary morphological representations, concurrently establishing global spatial awareness through positional embeddings. Finally, we introduce the Localized Spatial Attention (LSA) mechanism, which differentiates local interactions from broader semantic dependencies to adaptively construct local spatial awareness.

**Results**

SMM-MIL demonstrates competitive performance across all seven cohorts, surpassing the established methods by 4.6% in mean C-index, even when using only 10% of the patches per WSI for both training and inference. These results highlight the effectiveness of our collaboratively designed modules in delivering substantial improvements in model performance.

**Conclusion**

SMM-MIL offers a novel perspective for processing gigapixel WSIs. Thanks to its dual-feature extraction and the synergistic design of SMM-MIL. the proposed framework demonstrates strong potential.

# 26 - Living Metabolic Histopathology: A Prototype AI-Guided Real-Time Stain-Free 4D Digital Pathology Platform

**Kirtikumar Pilankar[1], Vikas Maindad[2]**
[1] MEDICHECK DIAGNOSTIC CENTER
[2] GRANT MEDICAL COLLEGE

Authors Dr. Kirtikumar Pilankar, Dr. Vikas Maindad

**Background**
Conventional histopathology relies on tissue fixation, staining, and static microscopic evaluation of non-viable tissue. Emerging advances in digital pathology, label-free optical imaging, metabolic imaging, and artificial intelligence (AI) have created the possibility of analyzing fresh living tissue in real time. “Living Metabolic Histopathology” is proposed as a futuristic diagnostic approach integrating viable tissue imaging, functional metabolic assessment, and computational pathology into a dynamic 4D diagnostic platform. Materials and Methods A conceptual prototype model named the “Living Metabolic Histopathology Platform (LMHP)” was designed. Fresh surgical or biopsy tissue is maintained in a sterile microfluidic live-tissue chamber with controlled oxygenation and temperature. The prototype integrates label-free imaging technologies including Quantitative Phase Imaging, Raman spectroscopy, autofluorescence imaging, hyperspectral imaging, and Optical Coherence Tomography. Metabolic assessment is performed using intrinsic cellular biomarkers such as NADH and FAD autofluorescence signals to evaluate mitochondrial activity and tissue metabolism. Acquired imaging data are processed through an AI-based computational pathology engine trained to assess nuclear morphology, cellular density, mitotic activity, tissue architecture, metabolic signatures, and temporal tissue changes. Sequential imaging generates a 4D digital pathology dataset.

**Results**
The proposed system theoretically enables rapid stain-free visualization of viable tissue while preserving cellular activity and metabolic behavior. AI-assisted analysis may provide real-time malignancy prediction, metabolic viability mapping, tumor margin assessment, hypoxia detection, and dynamic monitoring of tissue response. The prototype may reduce dependency on conventional staining and shorten diagnostic turnaround time during intraoperative and precision oncology applications.

**Conclusion**
Living Metabolic Histopathology represents a next-generation concept combining viable tissue imaging, metabolic analysis, AI-driven diagnostics, and temporal digital pathology into a real-time stain-free workflow. This prototype model may redefine pathology from static morphological assessment toward dynamic functional and metabolic tissue diagnostics, forming a future foundation for computational and precision digital pathology.

# 27 - HepaAgent: A Training-Free Multi-Scale Agent for Interpretable Whole-Slide Diagnosis of Hepatocellular Carcinoma

**Susu Luo[1,2], Linghan Cai[3,4,5], Wei Dong[6,7], Wenyi Jin[8,9], Dankai Liao[3], Yufeng Wu[3], Xiaoyi Chen[3], Leshui Wang[1], Xinyu Lu[1], Ziyue Wang[10], Yueming Jin[10], Kefei Shen[11], Ben Jin[12], Weiqin Li[13], Miaoxia He[12], Tianyi Zhang[3], Shangqing Lyu[3], Zeyu Liu[3], Hui Dong[6,14], Hui Liu[1]**

[1] The Third Department of Hepatic Surgery, Shanghai Eastern Hepatobiliary Surgery Hospital
[2] Moores Cancer Center, University of California San Diego
[3] PuzzleLogic Pte Ltd
[4] School of Computer Science and Technology, Harbin Institute of Technology (Shenzhen)
[5] Else Kroener Fresenius Center for Digital Health, Faculty of Medicine, TUD Dresden University of Technology
[6] Department of Pathology, Shanghai Eastern Hepatobiliary Surgery Hospital
[7] Vincent Mary School of Science and Technology, Assumption University of Thailand
[8] Department of Biomedical Sciences, City University of Hong Kong
[9] Department of Clinical Laboratory, Renmin Hospital of Wuhan University
[10] Department of Electrical and Computer Engineering, National University of Singapore
[11] Beijing Neurosurgical Institute, Capital Medical University
[12] Department of Pathology, Changzheng Hospital, Naval Medical University
[13] Department of Pathology, Changhai Hospital, Naval Medical University
[14] Clinical Research Institute, Shanghai Eastern Hepatobiliary Surgery Hospital

**Background/Objective:**
Clinical translation of artificial intelligence in digital pathology remains limited by opaque models that fail to replicate the hierarchical, evidence-guided reasoning of expert pathologists. Hepatocellular carcinoma (HCC) diagnosis depends on integrating low-magnification architectural patterns with high-resolution cytological features, a process not explicitly modeled by existing deep learning or multimodal large language models (MLLMs). We aimed to develop and evaluate HepaAgent, a training-free agentic framework that reconceptualizes whole-slide image (WSI) interpretation as a guideline-anchored, multi-scale diagnostic workflow.

**Methods:**
HepaAgent links visual perception to an explicit knowledge base derived from authoritative clinical taxonomies, including the WHO classification and the Edmondson-Steiner grading system, encoded as a 45-task checklist across three hierarchical categories. The system performs adaptive zoom-in screening at 2.5×, 10×, and 40× magnifications, with an Agentic Analyzer triggering deeper inspection only when low-resolution evidence is insufficient. A Summarizer integrates multi-scale evidence into structured diagnostic reports with traceable evidence chains. We benchmarked HepaAgent against five state-of-the-art MLLMs on 1,210 HCC WSIs (3,158 closed-ended and 542 open-ended questions). A prospective reader study compared diagnostic accuracy and reading time between unassisted and HepaAgent-assisted pathologists. Zero-shot generalization was assessed on 50 combined hepatocellular-cholangiocarcinoma (cHCC-CCA) cases.

**Results:**
HepaAgent consistently outperformed both generalist and pathology-specific MLLMs across all diagnostic categories, with superior BLEU, ROUGE-L, and BERTScore on free-text reporting and substantially reduced hallucination rates. Pathologists assisted by HepaAgent achieved significantly higher diagnostic accuracy and shorter reading times than unassisted review (Mann-Whitney U test, $P<0.05$). Without retraining, HepaAgent generalized to cHCC-CCA, reaching 80.37% accuracy in microscopic description.

**Conclusion:**
HepaAgent provides an interpretable, guideline-anchored paradigm for human-AI collaboration in computational pathology, supporting transparent, auditable, and clinically aligned WSI interpretation.

## 28 - AI-Assisted Application Practice and Enhancement in Pathology

**Qing Li, Jing Hu, Ting Lei, Shaojie Sheng, Huimin Song**
The First People's Hospital Of Changzhou (The Third Affiliated Hospital of Soochow University)

**Objective:**
To explore the application and quality improvement of AI in routine pathology work based on AI-assisted diagnosis practice in our hospital since May 2024, and to accumulate experience for smart pathology development.

**Methods:**
The PathOrchestra pan-pathology large model was introduced, implementing 12 diagnostic modules including cervical cytology, gastrointestinal biopsy, metastatic lymph node carcinoma, prostate biopsy, and breast cancer IHC interpretation. Four major platforms were established: AI diagnosis, slide quality control, telepathology, and large model training and annotation.

**Results:**
Over 100,000 slides have been scanned, forming a workflow of "AI preliminary diagnosis and physician review". AI diagnostic accuracy continues to improve, with new modules such as lung cancer PD-L1, serous effusion cytology, and lung cancer diagnosis being progressively introduced.

**Conclusions:**
1. AI effectively enhances the efficiency and quality of pathological diagnosis, particularly in high-volume, time-consuming, and relatively straightforward cases such as prostate biopsy and cervical cytology, where AI significantly accelerates the workflow. 2. Beyond technical training, the conceptual understanding of AI usage needs continuous reinforcement. 3. Departmental workflows should be adjusted accordingly. 4. A feedback mechanism should be established to facilitate collaboration between diagnostic teams and engineers for continuous model optimization, enabling human-machine collaborative symbiosis.

## 29 - Improvement of AI Pan-Model in Cervical Cytology Diagnosis

**Qing Li, Jing Hu, Bo Tian, Lei Peng, Yang He**
The First People's Hospital Of Changzhou (The Third Affiliated Hospital of Soochow University)

**Objective:**
To explore the improvement in AI diagnostic accuracy based on the practice of over 50,000 cases of AI-assisted cervical cytology in our hospital from September 2024 to the present.

**Methods:**
The PathOrchestra pan-pathology model was introduced, implementing 12 diagnostic modules including cervical cytology, gastrointestinal endoscopic biopsy, and prostate biopsy, which were gradually applied within the department.

**Results:**
A workflow of "AI preliminary diagnosis and physician review" was established, improving the efficiency of cervical cytology by 30%. The initial diagnostic sensitivity was 83% and specificity was 68%, with poor performance in diagnosing fungal elements. During routine operation, weekly case reviews were conducted, and AI engineers optimized the diagnostic model. After five iterative model updates, the diagnostic sensitivity reached 99%, specificity reached 77%, and the model demonstrated excellent capability in identifying fungal elements.

**Conclusions:**
1. AI effectively improves the efficiency and quality of cervical cytology diagnosis. 2. The pathology model must be adapted to the department's staining, sectioning, and other routine practices, requiring collaboration between diagnostic physicians and AI engineers for model optimization. 3. Using the model's training and annotation platform, the development of modules for thyroid fine-needle aspiration cytology and serous effusion cytology has shown initial success.

## 30 - Integrated Spatial Multi-Omics Imaging and LCM-Proteomics Uncover Inflammatory and Fibrotic Niches in Pulmonary Fibrosis

**Esther Cheow, U-Ming Lim, Matthew Choo, Nikhil Tulsian, Aaron Zefrin Fernandis**
Msd International Gmbh (singapore)

Pulmonary fibrosis (PF) is a progressive and fatal lung disease characterized by irreversible tissue scarring, yet the spatially resolved molecular mechanisms driving disease progression remain incompletely understood. To address this gap, we developed an integrated spatial multi omics workflow combining sequential matrix assisted laser desorption/ionization mass spectrometry imaging (MALDI MSI), laser capture microdissection (LCM), and LC–MS/MS to interrogate metabolites, lipids, and proteins within the same lung tissue section. Fresh frozen PF lung sections were analyzed using high resolution MALDI MSI to map untargeted metabolites and lipids alongside targeted protein imaging, followed by MSI guided ROI selection and LCM for localized proteomic profiling. This multi omics approach revealed spatially dysregulated inflammatory mediators, including lysophosphatidic acid and arachidonic acid, which co localized with $CD68^{+}$ macrophage rich regions, suggesting a link between inflammatory metabolism and immune cell activity in PF. Fibrotic foci were precisely identified using α smooth muscle actin MALDI IHC and validated by pentachrome staining prior to proteomic analysis. Preliminary LC–MS/MS data demonstrated upregulation of fibrosis associated pathways, including TGF β signaling, within fibrotic niches. By preserving spatial context across molecular layers, this workflow provides a granular view of fibrotic microenvironments and enables discovery of inter omic relationships that are obscured in bulk analyses, supporting more informed therapeutic target discovery in pulmonary fibrosis.

# 34 - EVALUATING A DIGITAL PATHOLOGY WORKFLOW THAT ADDRESSES EXISTING PATHOLOGY DIAGNOSTIC DEFICIENCIES

**Xavier Farré[1], Joshua Kibera[2], Maureen Waithaka[3]**
[1] Public Health Agency Of Catalonia
[2] The Pathology Network
[3] The Pathology Network

**Background/Objective:**
Digital pathology offers a promising strategy to address diagnostic gaps in many low- and middle-income countries (LMICs), where access to surgical pathology services remains limited due to shortages of pathologists and inadequate laboratory infrastructure. Although many of the necessary components for a digital pathology ecosystem exist, they are often fragmented across public and private institutions, limiting efficient utilization. The objective is to evaluate the feasibility and impact of a centrally coordinated modular digital pathology network in improving access, turnaround time, and diagnostic quality of surgical pathology services across Kenya.

**Methods:**
We evaluate The Pathology Network, a centrally coordinated, modular digital pathology workflow designed to optimize fragmented institutional resources across Kenya. The workflow divides the surgical pathology process into interconnected operational modules, including tissue procurement, sample transportation, histology processing, slide dispatch, diagnostic reporting, and data archiving. Participating institutions are assigned defined roles within the network. A centralized software platform coordinates and integrates all workflow components, enabling comprehensive oversight and communication between institutions.

**Results:**
During the pilot phase, the workflow was implemented across a network of 30 hospitals in Kenya, allowing participating facilities to request diagnostic services and receive results from centralized laboratories and remotely located pathologists. Approximately 5,000 specimens are processed annually by three laboratories supported by 14 remote pathologists, spanning an average inter-institutional distance of 150 km. Initial findings demonstrate a 70–90% reduction in diagnostic turnaround times. Diagnostic quality also improved through the implementation of standardized synoptic reporting templates aligned with ICCR and CAP cancer protocols.

**Conclusion:**
A modular digital pathology workflow represents a practical and scalable approach to improving diagnostic efficiency and quality in LMICs through better utilization of existing resources. Further studies are needed to evaluate the long-term scalability, sustainability, and generalizability of this model.

## 37 - A!PathCourse: An AI-Driven Digital Pathology Education Platform Leveraging Vision–Language Models and Structured Annotation Workflows

**Kokhaur ONG, Yiyi LU, Shuyie BEH, Junaun TAN, Longjun LIN, Weimiao YU**
A*star Bioinformatics Institute (bii)

**Background / Objective:**
Gastric cancer remains a major global health challenge with high mortality and persistent limitations in pathology education, including fragmented training resources, lack of standardized datasets, and limited access to scalable expert instruction. Traditional pathology training is heavily dependent on senior pathologists and is time-intensive, creating bottlenecks in knowledge transfer. To address these gaps, we developed A!PathCourse, an AI-augmented educational platform built on the A!HistoClouds infrastructure for gastric tissue annotation. The objective is to establish a standardized and scalable medical education ecosystem that enhances diagnostic learning and accelerates pathology training through integrated vision-text AI systems.

**Methods:**
A!PathCourse is implemented using a full-stack architecture (Java, Spring Boot, Vaadin, MySQL) and integrates multimodal AI components. A locally deployed Qwen-based LLM supports syllabus generation, course structuring, and automated teaching summaries. A cloud-based vision-language model (SmolVLM-500M-Instruct) enables ROI-level image interpretation and pathology description generation. The Histo-LAB-AI module converts annotated tissue regions into structured diagnostic explanations, enabling explainable AI-driven learning. A key innovation is the ROI-to-text pipeline, where pathology image regions are transformed into editable educational content, supporting continuous human-AI refinement and scalable dataset expansion.

**Results:**
The platform enables senior pathologists to efficiently generate standardized teaching content, while students receive AI-assisted interpretive feedback on pathology images. The integration of vision-text LLMs reduces manual annotation burden and supports large-scale creation of high-quality educational datasets. The system also provides structured performance tracking, automated course generation, and streamlined workflow management, improving consistency and reducing administrative workload.

**Conclusion:**
A!PathCourse demonstrates the transformative role of AI-driven vision-language models in pathology education, enabling scalable knowledge transfer, efficient diagnostic training, and high-quality dataset creation. By integrating multimodal LLMs with structured annotation workflows, the platform bridges clinical expertise and machine intelligence, laying the foundation for explainable AI in gastric oncology education and next-generation pathology training.

## 38 - Clinical Robustness: Rethinking Foundation Models for Diagnostic Gleason Grading

**Xinmi Huo[1,2], Eric Monzon[1], Kok Haur Ong[1], Haoda Lu[1], Chengxiang Yuan[1], Weimiao Yu[1,3,4,5], Soo Yong Tan[2]**

[1] Intelligent Digital and Molecular Pathology Lab, Bioinformatics Institute, A*STAR
[2] Yong Loo Lin School of Medicine, National University of Singapore
[3] Institute of Molecule and Cell Biology, A*STAR
[4] School of Biological Science, Nanyang Technological University
[5] Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore

**Background / Objective:**
Gleason grading is fundamental to prostate cancer diagnosis, yet manual assessment is limited by subjectivity, inter-observer variability and time constraints. Real-world deployment of AI-based grading systems is further challenged by variations in staining, tissue preparation and whole-slide scanner platforms. Although task-specific computational pathology models have achieved strong performance, the rapid emergence of large-scale pathology foundation models raises an important question: whether general-purpose representations can match purpose-built models under clinically relevant variability. This study aimed to systematically compare optimized task-specific models and foundation model-based approaches for automated Gleason grading.

**Methods:**
We conducted a systematic evaluation using the Automated Gleason Grading Challenge 2022 (AGGC2022) dataset, comprising H&E-stained whole-slide images from prostatectomy and biopsy specimens, with a subset scanned across six devices. All images were curated with expert pixel-level tissue annotations. We compared highly optimized task-specific models, including the top six AGGC2022 challenge winners and our in-house A!Prostate model, with four leading pathology foundation models: UNI, prov-gigapath, Virchow and Phikon-v2. For foundation models, frozen patch-level embeddings were extracted and evaluated using linear classifiers.

**Results:**
Task-specific models consistently outperformed foundation model-based linear classifiers across all evaluation settings. The performance gap was most pronounced under scanner-induced domain shift, highlighting the difficulty of maintaining robust grading performance under real-world acquisition variability. Although foundation models showed general representation capacity, frozen embeddings with linear probing were insufficient to match optimized task-specific models in this clinically demanding setting.

**Conclusion:**
Task-specific model development remains essential for robust AI-assisted Gleason grading, particularly under scanner-related domain shift. While foundation models provide promising general-purpose representations, their clinical potential may require stronger integration with task-specific supervision, domain adaptation and end-to-end optimization rather than simple frozen-feature linear probing.

## 39 - Systematic Analysis of Foundation Models for Prostate Cancer Pathology

**Yiran Liu**
Nanyang Technological University

Based on whole-slide images, the grading of prostate cancer plays a key role in clinical diagnosis and treatment decisions. However, the traditional pathological assessment process is time-consuming and susceptible, especially when distinguishing similar Gleason patterns. Recently, the pathological foundation models have demonstrated strong representation learning capabilities through large-scale unsupervised pre-training, providing new opportunities for computational pathology applications. In this study, we conducted a systematic evaluation of two advanced pathological foundation models, UNI and GigaPath, to achieve automated processing of the International Society of Urological Pathology (ISUP) grading of prostate cancer. We investigated two different learning processes. Method 1 employed supervised hierarchical graph classification, using Gleason labels with guidance annotations, followed by hierarchical graph summarization. Method 2 adopted a weakly supervised framework, using attention-based multi-instance learning (ABMIL), where slide-level predictions were directly achieved through aggregation of slide-layer embeddings without explicit annotations. WSIs were divided into 256×256 tissue regions. Tissue filtering and annotation overlap strategies were applied during the preprocessing. The backbone network of the frozen foundation model was used for embedding extraction, followed by ISUP prediction through a classification layer. The results showed that the weakly supervised ABMIL framework outperformed the supervised patch-level pipeline in all evaluation metrics. The combination of UNI and Method 2 achieved the highest overall accuracy (49.09%), weighted F1, and micro AUC value, while the combination of GigaPath and Method 2 achieved the highest quadratic weighted kappa coefficient, indicating stronger consistency with the pathologicalist grading sequence. Confusion matrix analysis further demonstrated that slice-level aggregation improved the ability to distinguish higher ISUP grades and reduced the prediction collapse towards intermediate categories. Overall, the research results indicate that the pathological foundation models benefit significantly from the slide-level aggregation strategy. This study highlights the potential of the weakly supervised foundation model pipeline in robust and scalable prostate cancer grading.

# 40 - Designing an End-to-End Artificial Intelligence (AI) Deep Learning Pipeline for Hepatic Steatosis Quantification: Less is More?

**Nagalakshmi Jegannathan[1], Xiaoman Zhang[2], Jia Xuan Seow[1], Menghan Zhou[2], Seow Ye Heng[1], Tony De Rong Ng[3], Rick Mong Goh Siow[2], Huazhu Fu[2], Yong Liu[2], Lionel Tim-Ee Cheng[4], Chee Leong Cheng[1], Tony Kiat Hon Lim[1], Li Yan Khor[1], Guo Lin Goh[5], Wei-Qiang Leow[1]**

[1] Singapore General Hospital
[2] Institute of High Performance Computing, Agency for Science, Technology and Research
[3] Ministry of Health Holdings
[4] Department of Diagnostic Radiology
[5] School of Biological Sciences, Nanyang Technological University

**Background:**
The global prevalence of metabolic dysfunction-associated steatotic liver disease (MASLD) is around 30% and expected to increase over time. Assessment of hepatic steatosis (HS) is currently graded in a semi-quantitative manner by pathologists, and is subject to inter- and intra-observer variability. We describe the evolution of a deep learning pipeline, utilising different model architectures such as U-NET and MedSAM, in HS quantification in patients with MASLD.

**Methods:**
Whole slide images (WSIs) from 41 MASLD patients' archival liver core biopsies between 2005-to-2017 were used. The initial model was trained and validated on 12154 annotated ROIs, and was based on U-NET. In our refinement phase as well as to overcome our small dataset limitation, we employed MedSAM, a pre-trained foundation model and a 5-fold cross validation step, into our pipeline. Both models calculated fat percentage and translated them into steatotic grades. We compared the performance of both models against expert pathologists' grades.

**Results:**
The weighted Kappa value for U-NET was 0.669 and 0.610 for MedSAM when compared against expert pathologists' steatosis grading. The DICE score for HS was 0.8955 for U-NET and 0.8542 for MedSAM. The F1 score for U-NET was 0.657 and 0.864 for MedSAM. The AUROC score for U-NET was 0.785 and 0.898 for MedSAM.

**Conclusions:**
We developed and refined an end-to-end AI deep learning pipeline that shows substantial agreement with expert pathologists. In our refinement phase, we employed MedSAM in order to reduce the need for large volumes of annotated pathology data. Although the MedSAM model achieved better F1 and AUROC scores, the DICE score and weighted kappa values were lower. We conclude that using a pre-trained foundation model such as MedSAM can overcome the limitation of a small dataset.

# 42 - Digital pathology(DP) validation for primary diagnosis in a tertiary care cancer centre

**Anila Sharma[1], Gurudutt Gupta[1], Meenakshi kamboj[1], Ankur Kumar[2], Anurag Mehta[1]**
[1] Rgcirc
[2] Manipal hospital

**Background:**
Digital pathology (DP) for primary diagnosis requires rigorous validation, particularly in resource-limited settings where international guidelines may not be directly applicable.

**Objective:**
To validate whole slide imaging (WSI) for primary diagnosis in a tertiary care cancer Centre and assess concordance with conventional microscopy.

**Methods:**
A two-phase validation study was conducted in accordance with NABL and international guidelines. Phase 1 included 60 prospective biopsy cases with a minimum 2-week washout period. Phase 2 involved real-time reporting of routine cases for two months, using both digital and glass slides. Slides were scanned at 40× magnification. Three pathologists participated, and discrepancies were reviewed and resolved. In phase 3, two pathologists assessed the validation of digital pathology using 3 months of cases (i.e., >100 cases) and performed Likert scoring to assess the ease of digital reporting. Pathologist 1 had more than 25 years of experience in histopathology reporting, while pathologist 2 had 15 years of experience in reporting.

**Results:**
All pathologists achieved ≥95% concordance between digital and glass slide diagnoses. Discordant cases were analysed to improve diagnostic agreement. Key challenges included scanner limitations, lack of full system integration, dependence on slide quality, and display constraints. The mean Likert score of pathologist 1 was 4.62 & slide referral rate was 20%. The Likert score for pathologist 2 was 4.47, and the slide referral rate was 26.5%.

**Conclusion:**
DP is non-inferior to conventional microscopy for primary diagnosis when validated systematically. A phased validation approach improves pathologists' confidence and ensures safe adoption. Hybrid reporting and workflow optimisation can address infrastructural challenges, facilitating DP implementation in resource-constrained settings.
Keywords: Digital pathology, validation, whole slide imaging, primary diagnosis, concordance

## 44 - Manga meets microscopy: Gemini-generated visual narratives to improve medical students' understanding of digestive system pathology

**Thiyaphat Laohawetwanit, Sompon Apornvirat**
Division of Pathology, Chulabhorn International College of Medicine, Thammasat University

**Background / Objective:**
Although multimodal large language models perform well in text-based applications, their ability to interpret pathology images remains limited. Recent advances in Gemini (Google, CA, USA) have introduced image-generation features that may support pathology education. This study evaluated whether Gemini-generated manga-style visual narratives could improve medical students' understanding of digestive system pathology.

**Methods:**
Gemini 3 Pro Image, released in late 2025, was used to generate manga-style educational visual narratives for 30 digestive system diseases using a two-step index prompt: (1) generation of a manga script and (2) generation of manga illustrations based on the script. Each script contained six panels covering the definition, etiology/pathogenesis, morphologic features, clinical issues, and a brief conclusion. Prompts emphasized high-yield concepts, simple language, minimal text, and clear educational messages. Two pathologists reviewed the generated content, and inaccuracies were corrected through follow-up prompts. Sixty medical students were divided into two groups (with and without manga-based visual narratives) and received identical lectures and study materials. Performance on 30 multiple-choice questions assessing understanding of digestive system diseases was compared between the groups.

**Results:**
Thirty manga-style educational visual narratives were generated using Gemini 3 Pro Image. Minor inaccuracies were identified in 11 (36.7%) narratives, including redundant text, inaccurate illustrations, obsolete concepts, missing treatment options, and misspellings; no major inaccuracies were observed. These issues were readily corrected through follow-up prompts. Students who received manga-based visual narratives achieved significantly higher scores than those who did not (mean score: 25.1 [83.7%] vs. 22.7 [75.7%], $p < 0.01$). Students also reported that these tools were helpful for studying pathology and expressed interest in learning prompts for generating educational manga.

**Conclusion:**
Gemini-generated manga-style visual narratives may serve as a practical adjunctive educational tool in pathology. Although minor inaccuracies may occur, they can be readily corrected through expert review and follow-up prompts.

## 46 - A!Secure: Security Infrastructure for Digital Pathology Ecosystems

**Longjun Lin, Kokhaur Ong, Jun Aun Tan, Weimiao Yu**
Agency For Science, Technology And Research (a*star)

**Background / Objective:**
Digital pathology systems manage whole-slide images, derived tiles, annotations, and metadata across viewers, storage services, and institutional users. Existing deployments often rely on storage-level encryption and network controls, while application-layer access may still depend on viewer logic, exposed storage paths, or client-supplied identifiers. This work develops an application-layer security framework for authorization, encryption, and auditability in digital pathology workflows.

**Methods:**
A!Secure was designed as a middleware layer between digital pathology applications and storage backends. It exposes functionality through a Software Development Kit, SDK, and Representational State Transfer(REST) API for integration with web applications and existing platforms. The framework combines backend-resolved identity, institution-aware authorization, audit logging, and envelope encryption. Image tiles and annotation objects are encrypted as independent resources and decrypted on demand through a streaming interface. For raw WSI protection, the framework implements compression-format-aware image encryption, where encryption is applied according to the internal structure of the underlying compressed image format rather than treating the WSI as an opaque file. JP2K was used as the initial proof-of-concept format.

**Results:**
A prototype supported encrypted DZI tile access, encrypted JSON annotation handling, and compression-format-aware WSI protection. In the JP2K proof-of-concept, protected compressed payload regions were encrypted while format-level structural information required for parsing and streaming was preserved. Tile decryption-on-read achieved approximately 30 ms average latency per tile. JP2K-based WSI encryption and decryption completed in 17.73 s and 6.04 s, respectively. Encrypted tissue regions showed strong visual distortion, with PSNR of 7.42 dB and SSIM of 0.0289. The framework was integrated into an existing digital pathology platform, demonstrating its compatibility with a web-based digital pathology workflow.

**Conclusion:**
A!Secure provides a deployable security layer for digital pathology systems by combining application-layer authorization, encrypted streaming access, audit logging, and compression-format-aware WSI protection without exposing storage locations or cryptographic keys to client applications.

# 48 - MMAP: A Multi-Magnification and Prototype-Aware Architecture for Predicting Spatial Gene Expression

**Dang Nguyen[1], Huy Pham Nguyen Dang[2], Duc Nguyen The Minh[1], Thai Nguyen Dac[1], Hang Thi Nguyen[3], Duong M. Nguyen[4]**

[1] Institute For AI Innovation And Societal Impact, Hanoi University Of Science And Technology
[2] Amsterdam High School for the Gifted
[3] College of Health Sciences, VinUniversity
[4] University of Illinois Urbana-Champaign

Spatial Transcriptomics (ST) enables the measurement of gene expression while preserving spatial information, offering critical insights into tissue architecture and disease pathology. Recent developments have explored the use of hematoxylin and eosin (H&E)-stained whole-slide images (WSIs) to predict transcriptome-wide gene expression profiles through deep neural networks. This task is commonly framed as a regression problem, where each input corresponds to a localized image patch extracted from the WSI. However, predicting spatial gene expression from histological images remains a challenging problem due to the significant modality gap between visual features and molecular signals. Recent studies have attempted to incorporate both local and global information into predictive models. Nevertheless, existing methods still suffer from two key limitations: (1) insufficient granularity in local feature extraction, and (2) inadequate coverage of global spatial context. In this work, we propose a novel framework, MMAP (Multi-MAgnification and Prototype-enhanced architecture), that addresses both challenges simultaneously. To enhance local feature granularity, MMAP leverages multi-magnification patch representations that capture fine-grained histological details. To improve global contextual understanding, it learns a set of latent prototype embeddings that serve as compact representations of slide-level information. Extensive experimental results demonstrate that MMAP consistently outperforms all existing state-of-the-art methods across multiple evaluation metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), and Pearson Correlation Coefficient (PCC).

# 49 - Exploring the Synergy of Morphometric Analysis and the Bethesda System in Thyroid Cytopathology in Category III, IV and V Lesions and Its Correlation with Histopathology: A Tertiary Care Experience in north east India.

**Mithlesh Bhargav[1], Mimna VM[2]**
[1] AIIMS Gorakhpur
[2] AIIMS Gorakhpur

**Background:**
Fine needle aspiration cytology (FNAC) reported using the Bethesda System for Reporting Thyroid Cytopathology (TBSRTC) is the cornerstone of pre-operative evaluation of thyroid nodules.However, indeterminate categories, particularly Categories III, IV, V remain diagnostically challenging because of overlapping cytomorphological features.Digital morphometric analysis provides objective quantification of nuclear and cellular characteristics and may improve diagnostic accuracy.

**Aim:**
To evaluate the utility of morphometric parameters in thyroid FNAC across Bethesda Categories II–V and correlate cytological findings with clinical, ultrasonographic and histopathological features. Materials and Methods: A cross-sectional observational study was conducted at a tertiary center. A total of 125 thyroid FNAC cases were analyzed: 20 Category II (benign controls) and 35 each from Categories III, IV and V. Morphometric parameters—nuclear area, nuclear perimeter, nucleus circularity, nucleus eccentricity, cell area, cell perimeter, cell circularity, cell eccentricity, and N/C ratio were measured on Papanicolaou-stained smears using QuPath (v0.4.3) digital pathology software. One-way ANOVA with post-hoc Tukey's HSD test was used for statistical comparison.Histopathological correlation was performed in all operated cases.

**Results:**
Females constituted 84.8% of cases, with peak incidence in 31–40year age group. Nuclear area increased progressively from 18.66 µm² in Category II to 58.30 µm² in Category V ($p<0.001$).The N/C ratio also showed significant escalation from 0.20 to 0.35 across Categories II–V ($p<0.001$). Nuclear circularity decreased with increasing Bethesda category, reflecting progressive nuclear irregularity. All nine morphometric parameters demonstrated statistically significant differences across categories. Histopathological malignancy rates were 0%, 8.57%, 11.43%, and 88.57% for Categories II, III, IV, and V, respectively. Papillary thyroid carcinoma was the predominant malignancy in Category V cases (90.3%). ROC analysis demonstrated high discriminatory performance for nuclear area (AUC 0.89) and N/C ratio (AUC 0.91).

**Conclusion:**
Digital morphometric analysis provides objective and reproducible parameters that complement conventional thyroid cytology. It may improve malignancy risk stratification, in indeterminate thyroid lesions.

# 50 - Audit of Repeat Immunohistochemistry Testing in a High-Volume Histopathology Laboratory: Root Cause Analysis and Quality Improvement

**GURUDUTT Gupta, Meenakshi Kamboj, Sangeeta Arora, Anila sharma, Sunil Pasricha, Anurag Mehta**
Rajiv Gandhi Cancer Institute & RC

**Introduction:**
Automated immunohistochemistry (IHC) staining has improved standardization and reproducibility in high-volume histopathology laboratories. However, repeat testing remains a challenge, leading to increased costs, delayed reporting, and potential loss of precious tissue, particularly in small biopsies and biomarker studies. The use of on-slide external controls is critical for ensuring assay reliability.

**Aim:**
To evaluate the frequency and underlying causes of repeat IHC testing and implement corrective strategies to reduce repeat rates.

**Methods:**
A prospective audit was conducted in a high-volume pathology department performing approximately 50,000 IHC tests annually. All repeat tests from January to December 2025 were recorded in a dedicated log. Data collected included antibody used, reason for repeat testing, platform, and corrective action. Root cause analysis (RCA) was performed for each case, and preventive measures were implemented. IHC staining was performed on fully automated platforms (Two different manufacturers) using both ready-to-use (RTU) and concentrated antibodies, with on-slide external controls in all cases.

**Results:**
The overall repeat rate was 1–2%. The most common cause was failure of on-slide control staining (67%), followed by primary antibody-related issues (15%), technician/pathologist errors (9%), detection system issues (4%), batch failures (3%), and pathologist-driven repeats (1%). Concentrated antibodies contributed more to repeat testing compared to RTU antibodies. Frequently repeated markers included CD30, PAX8, Ki-67, TTF-1, MLH1, PMS2, and PD-L1. Human factors, platform variability, and control tissue quality were key contributors.

**Conclusion:**
Automated IHC provides reliable and standardized results; however, repeat testing persists due to multifactorial causes. On-slide external controls are essential for quality assurance. Continuous audit, RCA, and targeted interventions can help achieve a repeat rate below 1%. Strengthening training, reagent quality monitoring, and platform-specific optimization are crucial for improving diagnostic accuracy and laboratory efficiency.

# 54 - ASSESSMENT OF SNOMED CT COVERAGE AND GAPS IN PROSTATE PATHOLOGY TERMINOLOGY FOR DIGITAL PATHOLOGY APPLICATIONS

**Xavier Farré[1], Maureen Waithaka[2], Joshua Kibera[3], Lorena Estévez-Iglesias[4], Hongmei Zeng[5], Paul Seegers[6]**

[1] Public Health Agency Of Catalonia
[2] The Pathology Network
[3] The Pathology Network
[4] University of León
[5] National Cancer Center / National Clinical Research Center for Cancer / Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College
[6] Palga, National Pathology Databank

**Background/Objective:**
Digital pathology increasingly relies on standardized terminologies to enable semantic interoperability, data sharing, and computational analysis. SNOMED CT is a comprehensive clinical terminology widely adopted in healthcare, yet its coverage and granularity for prostate pathology remain poorly understood. The objective is to review SNOMED CT concepts relevant to prostate pathology, assess their semantic adequacy, and identify gaps that affect diagnostic reporting, research, and digital workflows. This becomes even more important in worldwide collaboration.

**Methods:**
A literature review and terminology audit were conducted. Prostate pathology entities, including benign conditions, premalignant lesions, adenocarcinoma subtypes, grading, staging, biomarkers, and procedures, were mapped to SNOMED CT. Concept coverage, hierarchy, synonyms, and postcoordination patterns were evaluated.

**Results:**
SNOMED CT demonstrated broad coverage of common prostate diagnoses and procedures, with a robust hierarchical structure. However, inconsistencies were observed in grading systems, including Gleason patterns, variant histologies, and biomarker representation. Limited precoordinated terms necessitated complex postcoordination, potentially reducing usability and interoperability. Synonym variability and ambiguous concept definitions introduced a risk of coding discordance. Gaps were identified for emerging entities and molecular features. Limitations included evolving releases and potential regional extensions, which may affect generalizability.

**Conclusion:**
SNOMED CT provides a strong foundation for prostate pathology terminology but requires targeted refinement to support precise digital pathology applications. Enhancements to grading representation, variant classification, and biomarker encoding, along with clearer definitions, with SNOMED CT Terminology Bindings, and expanded synonyms, would improve consistency. Collaborative curation between terminologists and pathologists is recommended to optimize clinical reporting, data integration, and secondary use of prostate pathology data within interoperable healthcare systems.

## 57 - GLANCE: Gigapixel Level-wise Adaptive Navigation with Controlled Examination

**Smily Bharadwaj**
Indian Institutes Of Science Education And Research Bhopal

**Background:**
Whole-slide image (WSI) prediction under weak labels is bottlenecked by static feedforward architectures. All WSI methods expend identical compute regardless of difficulty, with no iterative refinement or learned termination. They cannot spend more compute on ambiguous cases nor stop early on obvious ones. This contrasts with LLMs, where test-time compute scaling improves reasoning. The question we address is: can WSI analysis benefit from adaptive compute, enabling a model to "look closer" when uncertain and "stop early" when confident?

**Method:**
We introduce GLANCE, a framework that reformulates slide-level prediction as a partially-observed sequential decision process over a frozen pathology foundation model. A learned policy adaptively selects (location, magnification) queries; a recurrent state aggregates evidence; and a learned termination signal decides when to predict. GLANCE is trained in three stages: behavioral cloning from an attention-MIL teacher to avoid a cold start over a large action space, trajectory-consistency self-supervision on 30k unlabeled TCGA slides, and outcome-aware policy optimization (GRPO) with a KL anchor to preserve consistency. Inference exposes a deployment-time compute budget knob that sweeps the accuracy-compute Pareto curve without retraining.

**Results:**
On CAMELYON16,TCGA-NSCLC, TCGA-RCC,PANDA, BRACS, and two OOD tests (CAMELYON17,CPTAC), GLANCE outperforms all constant-compute baselines. It achieves 0.987 AUC on CAMELYON16 (+1.4) and 0.872 κ on PANDA(+3.4). Stage B self-supervision drives most gains (+2–4 points), reducing OOD AUC drop from 0.082 to 0.038. On small-data BRACS, GLANCE improves by 4.7 points (0.818 vs. 0.771). GLANCE dominates the accuracy–compute Pareto frontier: at low budget, it reaches 0.912–0.951 AUC; at full budget, 0.987 AUC.

**Conclusion:**
Adaptive sequential examination with frozen foundation models enables-for the first time in WSI analysis-test-time compute scaling that is conditioned on input difficulty. GLANCE replaces fixed-cost weak supervision with a deployable, budget-aware framework. We propose the accuracy–compute Pareto frontier as the new evaluation standard for future adaptive WSI methods.

# 60 - Optimization of AI-Driven Universal Adversarial Perturbations for Intellectual Property Protection in Digital Pathology Images

**Kun Seng Vu[1], Longjun Lin[2], Kokhaur Ong[2], Eric Monzon[2], Weimiao Yu[2]**
[1] Nanyang Technological University
[2] Agency for Science, Technology and Research (A*STAR)

**Background/Objective**
The increasing release of gigapixel H&E whole-slide images (WSIs) for clinical and research purposes introduces significant data governance risks, where pathology images may be covertly exploited for unauthorized AI model training or inference without consent from the originating institution. Existing protection mechanisms, including access control and watermarking, are insufficient because released images are shared beyond authorized control, while visible perturbations compromise diagnostic quality. This study proposes a clean-label adversarial defense framework designed to protect pathology image datasets against unauthorized segmentation model usage while preserving visual fidelity.

**Methods**
The proposed approach embeds a universal adversarial perturbation (UAP) within a single $l_\infty$-bounded tile generated from stratified tissue regions. The perturbation is trained using surrogate segmentation models with the Deep Information Bottleneck (DIB) strategy to improve generalization across unseen architectures. A masked tiling approach enables seamless embedding into gigapixel WSIs of arbitrary size without requiring per-image optimization, thereby reducing deployment cost. The same adversarial texture, trained using DeepLabV3-ResNet101 and SegFormer-B4 surrogates, is applied in both inference-time evasion and training-time poisoning experiments.

**Results**
The proposed defense demonstrated strong black-box effectiveness against unseen segmentation architectures. On a held-out HRNet-OCR victim model, the embedded perturbation reduced mean intersection-over-union (mIoU) performance by 13–17% on 3000×3000 image patches during inference. Furthermore, poisoning the training dataset of an unrelated FCN-ResNet101 model under clean-label conditions resulted in catastrophic degradation, causing an 86% reduction in segmentation performance. The consistent effectiveness across multiple unseen architectures demonstrates strong generalizability and robustness under realistic black-box threat scenarios.

**Conclusion**
This work presents an effective AI-driven defense mechanism for safeguarding pathology WSIs against unauthorized AI exploitation while preserving diagnostic image quality. By embedding adversarial security directly into pathology data, the framework complements legal and institutional protections and establishes a foundation for secure medical imaging governance across future digital pathology and healthcare AI ecosystems.

# 65 - Segmentation-Guided Deep Learning for Cervical Cell Classification in Liquid-Based Thin-Layer Cytology

**Jinxi Li*[1], Haoda Lu*[2], Xinmi Huo[2], Kok Haur ONG[2], Weimiao Yu†[1,2]**
[1] Nanyang Technological University
[2] Bioinformatics Institute, A*STAR

**Background / Objective:**
Cervical cancer remains an important global health concern, and automated analysis of cervical liquid-based thin-layer cell smears may improve the efficiency and consistency of cytology screening. This study aimed to develop and evaluate a deep learning pipeline for seven-class Bethesda System (TBS) classification, binary cell segmentation, and segmentation-enhanced classification on cervical cytology images.

**Methods:**
A classification dataset of 38,337 cervical cell images labelled with seven TBS categories was split at the whole-slide image level to reduce information leakage. ResNet50 and SwinV2-Tiny were evaluated for multi-class classification, with a learnable source embedding indicating whether each image was from a single-cell or multi-cell dataset. SegFormer was trained on 4,101 annotated images for binary foreground-background cell segmentation and compared with nnU-Net. Four segmentation-guided classification strategies were investigated using frozen SegFormer outputs: soft-mask input enhancement, feature-level gated fusion, cross-attention fusion, and cross-attention fusion with CAM alignment.

**Results:**
SwinV2-Tiny outperformed the ResNet50 baseline, achieving a test macro F1 of 0.6461, with the largest gains observed in morphologically ambiguous categories such as LSIL and ASC-H. SegFormer achieved high binary segmentation performance, with a test Dice of 0.9377, and outperformed nnU-Net. However, none of the segmentation-guided fusion strategies surpassed the standalone SwinV2-Tiny classifier in overall macro F1. Localized per-class improvements were observed in some morphologically ambiguous categories, such as ASC-US and ASC-H, suggesting that segmentation-derived spatial cues may provide category-specific benefits despite limited overall gains.

**Conclusion:**
SwinV2-Tiny provided a stronger classification baseline than ResNet50, whereas accurate binary cell segmentation did not improve overall TBS classification on pre-cropped cell images. Segmentation-derived spatial cues may still benefit specific morphologically ambiguous categories. Future work should explore finer-grained targets, such as nucleus-cytoplasm separation or instance-level segmentation, to provide more diagnostically relevant structural information.

# 67 - Cross-Modal Fusion for Multi-Label Chest X-Ray Diagnosis: Toward Differential Diagnosis Support

**Shi Jingtian[1], Weimiao Yu[1,2], Piumi Don Simonge[2]**

[1] School of Biological Sciences, Nanyang Technological University

[2] Bioinformatics Institute, Agency for Science, Technology, and Research

Medical image interpretation requires clinicians to simultaneously identify multiple potential and co-existing pathologies, a process known as differential diagnosis (DDx). However, current AI-based medical image analysis frameworks treat disease prediction as independent binary classification tasks, failing to capture the interdependent and multi-label nature of clinical diagnosis. In this study, we operationalise the DDx problem as a multi-label task, where the model estimates the probabilities of multiple possible radiological findings and ranks them as diagnostic candidates. Using 149,781 chest X-ray (CXR) images from the MIMIC-CXR-JPG v2.0.0 dataset, we conduct controlled experiments to study the effects of pretrained encoders, language models, loss functions, and multimodal fusion strategies. Models are first evaluated on 14 CheXpert labels, and the best configuration is further validated on 12 expert-curated disease labels using 5-fold patient-level cross-validation. The cross-modal architecture combines visual features from pretrained image encoders (BEITv2, DINOv2, RAD-DINO) with clinical indication text encoded by BERT and PubMedBERT, fused through lightweight multi-layer perceptron (MLP) modules. The best-performing configuration (RAD-DINO + PubMedBERT + MLP fusion with dual-head ranking) achieves a macro AUROC of 0.8077 ± 0.0034 and an F1-score of 0.602 ± 0.004 on 12 clinical disease labels under 5-fold cross-validation. Key findings include (1) domain-specific pretrained models (RAD-DINO) consistently outperform general-purpose models, (2) Asymmetric Loss (ASL) improves recall by 36% over Binary Cross-Entropy (BCE) and is beneficial in high-sensitivity settings, (3) simple MLP-based fusion matches cross-attention mechanisms while using only 1/22 of the parameters of the Bitransformer baseline. Overall, these results demonstrate that domain-specific pretraining, ASL, and efficient cross-modal fusion enable strong multi-label performance without architectural complexity. Importantly, this unified framework is broadly applicable across diverse medical imaging modalities beyond chest radiography, highlighting its potential for generalisable deployment in clinical AI systems.

# 72 - Developing a Multi-format Image Viewer for Multidisciplinary Team (Tumor Board) Discussions and Education

**Jayanthi Karunanithi[1], Malay Singh[3], Joanna Stępień[3,4], Jinglin Li[3,5], Yang Fu[3,6], Arvind Channarayapatna Srinivasa[3], Isabelle Jang[1,2], Constance Teo[1,2], Kotamma Venkateswaran[7], Seow Ye Heng[1], Hwee Kuan Lee[3,8,9], Li Yan Khor[1,2]**

[1] Singapore General Hospital
[2] Duke-NUS Medical School
[3] Bioinformatics Institute (BII), Agency for Science, Technology and Research (A*STAR)
[4] Faculty of Electrical Engineering Automatics Computer science and Biomedical Engineering, AGH University of Krakow
[5] College of Computing and Data Science, Nanyang Technological University
[6] Beijing Normal University
[7] Sengkang General Hospital
[8] International Research Laboratory on Artificial Intelligence
[9] A* Centre for Frontier AI Research (CFAR)

**Background/Objective**
This project aims to integrate 360-degree manipulation of the 3D gross specimen image, annotation of the object surface in space to highlight the location of tissue sections obtained for microscopic evaluation, and to co-register these sections with the corresponding whole slide image (WSI) within in a single web-based viewer. This integrated viewer was designed to support multidisciplinary discussions and facilitate teaching and learning.

**Methods**
Selected head and neck resection specimens from previously consented study patients were scanned with a multi-line laser scanner (Einscan Pro). The specimens were grossed as per normal clinical workflow. Glass slides were scanned to WSI (Philips SG300). RGB-format images of tissue sections, CT images and clinical reports were also collected. A system was built to consist of web-based frontend and backend services with databases for metadata, audit logs and object storage for large files (.obj, .TIFF, .jpg, .pdf and DICOM). Automated ingestion pipelines perform format conversion, quality control and indexing. Access control and audit trails ensure secure, reproducible use for both clinical review and educational activities.

**Results**
The viewer comprises five integrated modules (3D Model, Digital Pathology, CT Scan, Gross Image and Clinical Report) and can be viewed in various combinations. Users can rotate, zoom, and annotate 3D specimens, perform measurements on histology and gross images, explore CT volumes in multiple planes and review structured reports within the same interface. Annotations and edits are automatically versioned with case and user identifiers, ensuring full auditability. Cross-linking between specimen, imaging, and reports enables co-localization of findings across modalities. Export functions generate structured outputs and visual summaries suitable for tumor board minutes and teaching materials.

**Conclusion**
This viewer consolidates multiple image formats relevant to anatomic pathology workflow into a single interactive environment. This enhances multidisciplinary communication and provides a scalable framework for both clinical decision-making and medical education.

# 73 - Virtual EBER-ISH: Nucleus-level Pan-disease Epstein-Bar Virus Detection from Routine Hematoxylin-Eosin Staining Image

**Shengbing Zang[1], Shuangping Huang[2], Jingping Yun[1]**
[1] Sun Yat-sen University Cancer Center
[2] School of Electronic and Information Engineering, South China University of Technology, Guangzhou, China

**Background**
Epstein-Barr virus (EBV)-encoded RNA in situ hybridization (EBER-ISH) is the gold standard for detecting latent EBV in tissue samples, guiding diagnosis and management of EBV-associated malignancies. However, EBER-ISH is constrained by high costs, reliance on specialized molecular pathology infrastructure, and prolonged turnaround times, limiting its accessibility in low-resource and rural settings. We present virtual EBER-ISH (vEBER), a deep-learning framework using routine hematoxylin and eosin (H&E)-stained slides as a rapid, economical alternative.

**Methods**
We constructed PanEBV, the first nucleus-level EBV-status annotation dataset, encompassing approximately 340 million labeled nuclei across 15 disease entities from 27 medical centers. Based on this dataset, we developed a framework that jointly detects nuclei and infers EBV infection status from H&E images, incorporating a Disease-Infection Specialized Expert (DISE) module for clinically reliable calibrated confidence.

**Results**
vEBER achieved a mean nucleus-level area under the receiver operating characteristic curve (AUROC) of 0.952 across all disease entities, while DISE reduced the expected calibration error (ECE) by 78.2%. In a simulated clinical reading on an independent external cohort, vEBER improved pathologists' diagnostic accuracy from 68.0% to 83.3% without additional pathology procedures, while shortening turnaround time from days to minutes.

**Conclusion**
vEBER provides a cost-effective, time-preserving, and scalable solution for pan-disease EBV assessment.

# 75 - The Road Taken; From Our First MDT to AI Diagnostics for the GI Service at NHS's University Hospitals of Derby and Burton

**Mark Lloyd[2], Rahul Deb[1], Saeed Amal[3]**
[1] NHS's University Hospitals of Derby and Burton
[2] Fujifilm Healthcare
[3] Northeastern University

**Background:**
The Robert Frost poem The Road Not Taken reflects how choice is inevitable, but indecision can lead to regret. Digital Pathology (DP) is advancing rapidly, and we aim to share our experience from a large UK public hospital pathology service. We describe our success in establishing digital pathology for primary diagnosis and multidisciplinary tumor boards, and outline the road our GI service has taken to develop targeted digital use cases. This has expanded our experience as we begin collaborating on AI solutions. Common barriers to clinical DP implementation include change management, unfamiliarity, IT challenges, and internal advocacy. We will describe how UHDB embraced digital transformation and how we are leveraging this foundation to remain innovative into the future.

**Methods:**
We implemented and validated two Sysmex® P-1000 slide scanners for clinical sign-out. We launched our clinical digital workflow in 2025, followed shortly by our digital Breast MDT service. We have also begun developing GI-specific algorithms in partnership with Northeastern University, Boston, MA, USA. We will present the work completed to date and our initial findings.

**Results:**
Over six years, UHDB has invested in DP implementation. This has led to progress from early MDT use cases to emerging work in AI diagnostics. We will share the current state of DP at our organization with the ASDP audience, describing why we chose this road, the outcomes achieved, and what lies ahead.

**Conclusions:**
Many organizations continue to question how best to implement DP, while others are advancing the next stage of innovation. This oral presentation will share our experience and planned next steps at UHDB. We aim to provide a global perspective from a UK NHS site to help others take their own road toward digital transformation.

# 76 - Using an LLM-based AI Agentic Tool for Histopathology Case Compilation

**Yan Fei Ng[1], Navin Kumar[2], Daniel Kang[2], Soorya Ram Shimgekar[2], Piyum Zonooz[2], Chee Leong Cheng[1,3], Li Yan Khor[1,3]**

[1] Singapore General Hospital
[2] Nimblemind.ai
[3] Duke-NUS Medical School

**OBJECTIVE**

Pathology laboratory information system (LIS) searches are frequently performed to compile datasets for downstream applications such as research, audit studies, and education. Compiling datasets from legacy systems can be challenging due to proprietary database structures, including MUMPS-based dictionaries, which differ from modern relational databases. This study evaluated a large language model (LLM)-based AI agentic tool [Nimblemind Multi-agent System (nMAS)] for compiling a predefined histopathology dataset from unstructured free-text pathology reports.

**METHODS**

A set of 54 dummy gastric biopsy pathology reports was created based on reports from the Singapore General Hospital's Department of Anatomical Pathology. The reports represented diverse free-text reporting styles across multiple pathologists. The AI workflow (nMAS) used a modular multi-agent architecture consisting of document ingestion, report sectioning, context-aware retrieval, rule-based pattern matching, and LLM-driven feature extraction. Clinician-informed prompting, in-context learning (ICL), and grounding-aware validation were applied to ensure extracted outputs were explicitly supported by the source text. The system processed reports and identified specific features in the reports to compile a subset with Helicobacter pylori gastritis. The features included: Stomach, gastric, biopsy, positive for presence of Helicobacter pylori, Helicobacter pylori-like micro-organisms, and Helicobacter associated active chronic gastritis.

**RESULTS**

nMAS demonstrated high extraction performance: Detection accuracy was 100% for biopsy and gastric biopsy identification, 98.1% for Helicobacter pylori detection, and 96.3% for Helicobacter pylori-associated gastritis. Precision, recall, and F1-scores were ≥96% across all evaluated features.

**CONCLUSIONS**

An LLM-based AI agentic workflow (nMAS) can accurately compile structured histopathology datasets from heterogeneous free-text pathology reports, including reports derived from legacy LIS environments. The combination of multi-agent orchestration, context-aware extraction, clinician-informed prompting, ICL, and grounding validation enabled robust identification of clinically relevant gastric biopsy features with high precision and reproducibility. Further validation on larger and more complex datasets will help assess scalability and deployment in real-world pathology workflows.

# 77 - Infrastructure-Level Optimization of Whole Slide Imaging: Evaluation of a Performance Layer for Scalable Enterprise Pathology Access

**Yixiao Zhao[1], Val Anthony Alvero[1], Cheng Chee Leong[2], Kenneth Tang[1]**
[1] Precidx Corp.
[2] Singapore General Hospital

**Background / Objective:**
Multi-petabyte whole slide imaging (WSI) archives create operational challenges across clinical and pharmaceutical workflows. Native files are costly to maintain in high-performance storage, yet low-cost archival tiers introduce retrieval latencies, egress costs, and workflow friction for downstream analysis. This study evaluated an infrastructure-level performance layer that generates lightweight (approximately 80% reduction in data footprint), native-linked access representations from native WSIs. These representations are retained in high-speed storage tiers for instantaneous, scalable access and viewing, maintaining deterministic linkage to the authoritative native files secured in cold archives.

**Methods:**
One hundred native .iSyntax WSIs totaling 72.7 GB, digitized at Singapore General Hospital using a Philips scanner, were processed using the Precidx Optimization Platform (POP). POP generated lightweight, native-linked access representations in .pcdx files from the proprietary WSIs. These representations were evaluated for payload reduction, spatial structure, metadata preservation, viewer compatibility, and deterministic linkage to authoritative source files. Objective image fidelity was assessed using peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM), and Delta E metrics. Infrastructure performance was benchmarked using local baseline calibration to isolate transaction latency, modeling download efficiency, throughput virtualization gains, and network concurrency capacity.

**Results:**
The performance layer achieved a 78.17% payload reduction (15.8GB remaining) with high image fidelity (average PSNR: 44.70 dB, SSIM: 0.992, 95th percentile Delta E: 2.96). Benchmarking confirmed transaction latency parity (±1.5%). 100 Mbps network simulations demonstrated a 78.3% download efficiency gain, reducing retrieval from 1.66 hours to 21 minutes. This yielded a +358% effective throughput gain and expanded concurrent network capacity by 4.5x. The .pcdx files achieved interoperability within the ASAP open-source viewer.

**Conclusion:**
Transitioning to a WSI performance layer can resolve the financial and latency bottlenecks of cloud storage. The .pcdx files reduce time-to-first-view, generating storage and bandwidth egress savings for scalable, digital pathology.

# 80 - Resolving Isomorphic Heterogeneity in Nasopharyngeal Carcinoma: A Multimodal Computational Pathology Framework Integrating Histomorphology and Virtual Spatial Transcriptomics for Immunotherapy Prognostication

**Lina Hu[1], Yizhi Wang[2], Yonghong He[2], Jingping Yun[1]**
[1] Sun Yat-sen University Cancer Center
[2] Shenzhen International Graduate School, Tsinghua University

**Objective:**
Predicting immunotherapy outcomes in recurrent/metastatic nasopharyngeal carcinoma (R/M-NPC) is challenged by "isomorphic heterogeneity"—cryptic molecular divergence concealed within uniform morphological phenotypes on H&E slides. We aimed to develop a novel multimodal computational pathology framework integrating self-supervised histomorphological phenotype clusters (HPCs) with virtual spatial transcriptomics to resolve this discordance, decode microenvironmental complexity, and accurately stratify patient risk.

**Methods:**
In a pan-stage NPC cohort (n=3800), whole-slide image (WSI) features extracted via a foundation model (UNI) were clustered into HPCs. To resolve isomorphic heterogeneity, a 188-dimensional multi-omics feature space integrating UNI embeddings and 982 immune-related genes predicted via a generative spatial transcriptomics model (STPath) was constructed. Leiden clustering was applied to define high-resolution morpho-molecular phenotype clusters (MP-HPCs). A K-nearest neighbors classifier mapped these MP-HPCs across an immunotherapy cohort (n=968; 779 training, 189 validation). WSI-level MP-HPC area fractions underwent centered log-ratio (CLR) transformation to eliminate collinearity, followed by LASSO-Cox regression to develop a prognostic risk model.

**Results:**
Unimodal clustering of baseline morphological features delineated 35 HPCs. Multimodal integration of molecular features decoupled morphologically homogeneous regions into 53 granular MP-HPCs. Notably, nine HPCs decoupled into sub-clusters with diametrically opposed survival trajectories. These decoupled entities encompassed key microenvironmental compartments, including tumor nests (HPC-0/8), fibrous stroma (HPC-10/14/25), inflammatory infiltrates (HPC-5/15), and ciliated columnar epithelium (HPC-7/27). Histological re-evaluation confirmed nuanced architectural differences within five highly representative decoupled clusters. Clinically, the MP-HPC prognostic model outperformed the morphology-only HPC baseline, with the validation C-index improving from 0.655 (95% CI: 0.621-0.713) to 0.681 (95% CI: 0.634-0.743) in the combined R/M-NPC cohort.

**Conclusion:**
Fusing virtual spatial transcriptomics with self-supervised histomorphology effectively resolves isomorphic heterogeneity, providing a scalable and precise computational prognostic tool for R/M-NPC immunotherapy. Targeted spatial transcriptomic profiling of these decoupled clusters is currently ongoing to further elucidate their underlying biological mechanisms.

# 81 - Large Language Model-based System for Intraoperative Frozen Pathology Report Generation and Lymph Node Metastasis Prediction in Thyroid Cancer

**Xue Li[1], Shuangping Huang[2], Jingping Yun[1]**
[1] Sun Yat-sen University Cancer Center
[2] South China University of Technology

**Objective:**
Intraoperative frozen-section diagnosis of thyroid lesions is time-critical but routine reports are often brief and lack sufficient morphologic explanation. Existing report generation studies mainly focus on postoperative formalin-fixed, paraffin-embedded (FFPE) slides and do not address the domain gap to frozen sections. We developed a clinically interpretable framework for automated report generation from thyroid frozen sections, jointly supporting cancer diagnosis, morphologic description, and lymph node metastasis risk prediction.

**Methods:**
Scanned whole-slide images (WSIs) were first used for slide-preparation-adaptive fine-tuning with low-rank adaptation and domain-adversarial training to adapt a pathology foundation model from FFPE to frozen sections while preserving pretrained weights. The adapted morphologic representations were then organized into histomorphological phenotype clusters (HPCs), totaling 46 HPCs, annotated by pathologists, and integrated with diagnostic contribution analysis to identify diagnosis-relevant patches for report dataset construction and supervised training. Finally, a Transformer-based multiple-instance report generator was built on the adapted UNI visual encoder, incorporating diagnosis-guided key patch selection and supervised attention alignment to generate diagnostic conclusions with interpretable morphologic descriptions.

**Results:**
In 1,470 thyroid frozen-section cases evaluated by 5-fold cross-validation, slide-preparation-adaptive fine-tuning reduced the representation gap between FFPE and frozen sections, with improved overlap of feature embeddings after adaptation. The resulting HPCs showed strong pathologic semantic consistency and separated clinically meaningful morphologic patterns. Contribution analysis showed that papillary thyroid carcinoma- and nodular goiter-related clusters were dominant contributors to cancer diagnosis, whereas fibrosis- and inflammation-related clusters contributed most strongly to lymph node metastasis prediction, highlighting the importance of tumor microenvironmental patterns. Quantitatively, the framework achieved AUC/ACC/F1 of 0.961/0.948/0.930 for cancer diagnosis and 0.813/0.798/0.795 for metastasis risk prediction. Report generation achieved 93.52% diagnosis accuracy and 74.74% metastasis-risk accuracy.

**Conclusion:**
This framework provides clinically interpretable automated reporting for intraoperative thyroid frozen sections and may improve diagnostic performance and real-time surgical decision support.

## 82 - AI-Assisted Reduction of Unnecessary Immunohistochemistry Ordering in Nasopharyngeal Carcinoma Biopsy Evaluation Using Whole-Slide H&E Images

**Bingzhong Jing, Shengbing Zhang, Jingping Yun**
Sun Yat-sen Univerisity

**Objective:**
Immunohistochemistry (IHC) is typically indicated when H&E-based morphological assessment is equivocal during nasopharyngeal carcinoma (NPC) biopsy evaluation. However, junior pathologists often over-order IHC due to lower diagnostic confidence. We developed an artificial intelligence (AI) system leveraging whole-slide H&E images to reduce unnecessary IHC ordering while preserving diagnostic safety.

**Methods:**
The IHC-ordering system was trained on 1,313 retrospective H&E slides (832 NPC and 481 normal) from Sun Yat-sen University Cancer Center (SYSUCC). The system utilized an ensemble of six expert diagnostic models sharing a unified feature encoder to extract multi-dimensional descriptors, including diagnostic probability distributions, predictive uncertainty metrics, and morphology-aware quality-control features. These features were then integrated into an XGBoost-based IHC ordering model. The system was evaluated on a retrospective internal SYSUCC cohort (n=232), a Hunan external cohort (n=291), a Sichuan external cohort (n=295), and a prospective internal SYSUCC cohort (n=645).

**Results:**
Without any missed diagnoses (100% sensitivity across all cohorts), the AI system recommended IHC ordering in 18.53% of the retrospective internal SYSUCC cohort, 26.21% of the Hunan cohort, and 22.11% of the Sichuan cohort. Compared with clinicians, the AI significantly curbed unnecessary testing. In the retrospective internal SYSUCC cohort, the AI's recommendation aligned closely with senior pathologists (mean ordering rate: 18.32%), markedly lower than junior pathologists (69.64%). In the prospective internal SYSUCC cohort, the AI reduced the IHC ordering rate to 31.32%, yielding a 55% relative reduction compared with the actual routine physician ordering rate of 69.61%. SHAP analysis revealed that these selective ordering decisions were primarily driven by prediction uncertainty and morphology-related features.

**Conclusion:**
This AI-based IHC ordering system reduced unnecessary IHC use in NPC biopsy evaluation while preserving diagnostic safety across multi-center validation. This could assist in real-world clinical decision support.

## 84 - Diffusion-Based Data Augmentation for Artifact Classification in Digital Pathology

**Yingchen Shao[1], Shijun Qi[1], Longjie Li[2], Weimiao Yu[1,2]**
[1] Nanyang Technological University
[2] Agency for Science, Technology and Research

Digital pathology (DP) enables scalable image analysis but remains highly sensitive to image quality, as artifacts can significantly affect diagnostic outcomes and downstream task. Accurate identification and classification of artifact-containing images are therefore essential for reliable DP workflows, yet are often hindered by class imbalance and limited annotated data. In this study, we address this challenge by leveraging generative diffusion models for data augmentation in pathology image quality control. Specifically, we generate synthetic images and incorporate them into the training set to alleviate data scarcity and imbalance, thereby improving artifact classification performance. We combine weak and fine-grained annotation strategies and systematically explore multiple generation methods based on Stable Diffusion (SD), including Text-to-Image and DreamBooth, along with extensive parameter tuning. Experimental results demonstrate that synthetic data consistently enhances classification accuracy, with more pronounced improvements in low-data regimes. Specifically, accuracy increases from 80.8% to 83.6% (+2.8%) on a small-scale dataset (SD 3.5 with fine-grained annotations), from 95.0% to 95.6% (+0.6%) on a large-scale dataset (SD 1.5 with weak annotations), and from 73.7% to 78.6% (+4.9%) on a purely external dataset, highlighting the effectiveness of synthetic augmentation under varying data conditions. To further analyze the quality of generated data, we quantify the distribution gap between synthetic and real images using FID and KID, and visualize feature distributions via UMAP. The results show that synthetic images can effectively expand the data distribution boundary and partially compensate for underrepresented regions, while also improving coverage of rare artifact patterns. However, challenges such as distribution inconsistency and missing modes remain. Overall, a non-negligible distribution gap persists between synthetic and real data, indicating that synthetic images are better suited as a complementary resource rather than a replacement. When used appropriately, they can enhance model robustness and generalization in pathology image quality control tasks.

## 85 - A Curated Patch-Level Artifact Database and Robust Evaluation Framework for Digital Pathology Quality Control

**Shijun Qi[1,2], Yingchen Shao[1,2], Longjie Li[1], Weimiao Yu[1]**

[1] Intelligent Digital and Molecular Pathology (iDMP) Lab, Bioinformatics Institute (BII), Agency for Science, Technology and Research (A*STAR)

[2] Nanyang Technological University

Digital pathology is increasingly used to support clinical diagnosis, but its reliability depends on whether whole-slide images (WSIs) preserve diagnostically meaningful tissue morphology. In routine workflows, artifacts introduced during tissue preparation, staining, and scanning can distort morphological patterns and cause downstream AI models to learn artifact-related shortcuts rather than subtle histological structures. Therefore, effective upstream quality control and timely exclusion of artifact-contaminated regions are essential for reliable computational pathology analysis. However, fine-grained annotated artifact databases for pathology quality control remain limited. This study developed a comprehensive artifact-level database for H&E pathology image quality control. A total of 104,278 patches were extracted from more than 3,100 WSIs collected across multiple institutions and curated into a 12-class database through an iterative annotation and refinement process. The final database included ten common artifact categories, together with normal tissue and background. Using this database, five convolutional neural network backbones were evaluated for artifact classification. ConvNeXt-Tiny was selected as the main model because it provided the best overall balance between accuracy, robustness, and computational efficiency. To assess generalizability, the database was further divided into internal and external datasets, and diffusion-generated synthetic patches were added only to the training set. Real-synthetic training did not meaningfully improve internal performance, but increased external accuracy by 4.9%, from 73.7% to 78.6% at a 10% synthetic-data ratio. This suggests that generated data may be useful as a limited supplement, rather than a replacement, for real pathology images. Perturbation-based robustness testing further indicated that synthetic-data supplementation may improve external generalizability and model stability under image perturbations. Overall, this study provides a practical artifact-level QC workflow for digital pathology, highlighting the importance of careful data curation, efficient model selection, external validation, and robustness evaluation for reliable computational pathology analysis.

# 86 - EFFECT OF INTERNATIONAL SUBSPECIALTY CONSULTATION ON CHALLENGING GASTROINTESTINAL PATHOLOGY DIAGNOSES IN MONGOLIA

**Batzaya Nyambat[1], Jijgee Munkhdelger[1,2], Nyamdelger Enkhbold[1], Ichihara Shin[3], Oyun-Erdene Bat-Ochir[1]**

[1] Anatomic pathology department, Intermed IUHW hospital
[2] Department of pathology, Kameda Medical Center
[3] Department of diagnostic pathology, Asahikawa Medical University Hospital

**Background:**
Subspecialty consultations are vital for accurate pathological diagnosis; however, a lack of gastrointestinal (GI) subspecialists in Mongolia makes diagnosing complex conditions challenging. We aimed to assess the impact of international expert consultations.

**Methods:**
We retrospectively evaluated 49 GI cases sent from Intermed IUHW Hospital to a Japanese GI expert between 2022 and 2025 via email using digitized whole slide images (WSIs). Local preliminary diagnoses were compared with the expert's final diagnoses to calculate concordance.

**Results:**
Consultation cases involved small biopsies from liver (53.1%, n=26), pancreas (28.6%, n=14), stomach (6.1%, n=3), ampulla of Vater (4.1%, n=2), and esophagus, duodenum, ileum, and colon (2.0%, n=1 each). Out of 49 cases, 22 (44.9%) showed full concordance and 14 (28.6%) were concordant with comments, yielding 73.5% (n=36) agreement. Among cases concordant with comments, the most frequently noted comment suggested that intrahepatic cholestasis may be an indication of drug induced hepatic injury (DIHI) and recommended a Berlin-Blue stain to exclude iron deposition. Disease-specific analysis of the 36 agreed cases revealed the highest concordance occurred in primary biliary cholangitis (PBC) (11.1%, n=4), DIHI (8.3%, n=3), and PBC-AIH (autoimmune hepatitis) overlap, neuroendocrine tumor, and pancreatic ductal adenocarcinoma (5.6%, n=2 each). Diagnostic discrepancies occurred in 13 cases (26.5%), comprising 8 minor (16.3%) and 5 major (10.2%) discrepancies. The most frequent disagreements were neoplastic pancreatic disease and non-neoplastic liver disease (PBC vs AIH vs DIHI) among minor and major discrepancies, respectively. The mean turnaround time of the consultation was 20.35 ± 23.10 hours.

**Conclusion:**
Although our assessment demonstrated substantial concordance between preliminary and expert diagnoses, subspecialty consultation provided additional confidence in suspected diagnoses. Importantly, WSI-based international consultation significantly improves the diagnostic accuracy of complex gastrointestinal cases. It represents an invaluable, time-efficient approach that bridges critical subspecialty gaps and enhances the overall standard of patient care in Mongolia.

# 87 - Development and Local Validation of a ConvNeXt-Based Deep Learning Platform for Gleason Grading in Prostate Histopathology

**Chua Hsuan Kai[1], Chen Feng[1], Chan Wan Ling[2]**
[1] IMU University
[2] Sunway University

**Background:**
Gleason grading is essential for assessing prostate cancer aggressiveness and guiding treatment decisions. However, manual grading by pathologists is time-consuming and subject to significant inter-observer variability. o improve diagnostic consistency, we developed a deep learning–based platform for automated semantic segmentation and Gleason pattern classification in prostate histopathology images and validated its performance using a local hospital dataset.

**Methods:**
The model achieved an overall IoU of 0.8345 and accuracy of 0.9874. Recall for Gleason pattern 5 (G5) reached 0.9708, indicating high sensitivity for aggressive disease. F1 scores for Gleason patterns G3, G4, and G5 were 0.8405, 0.8216, and 0.8552, respectively. Slide-level validation demonstrated high concordance with pathologist diagnoses, particularly in high-grade tumours, although occasional discrepancies in pattern ordering (e.g., 5+4 vs. 4+5) were observed. Benign tissue segmentation remained comparatively challenging, with an IoU of 0.7261.

**Results:**
The model achieved an overall IoU of 0.8345 and accuracy of 0.9874. Recall for Gleason pattern 5 (G5) reached 0.9708, indicating high sensitivity for aggressive disease. F1 scores for Gleason patterns G3, G4, and G5 were 0.8405, 0.8216, and 0.8552, respectively. Slide-level validation demonstrated high concordance with pathologist diagnoses, particularly in high-grade tumours, although occasional discrepancies in primary and secondary pattern ordering (e.g., 5+4 vs. 4+5) were observed. Benign tissue segmentation remained comparatively challenging, with an IoU of 0.7261.

**Conclusion:**
The ConvNeXt-based platform demonstrated accurate segmentation and reliable Gleason grading on a local Malaysian dataset, supporting its potential as a clinical adjunct for standardized digital pathology workflows. Larger-scale validation studies are ongoing.

# 88 - PHH3-Guided Multiplex Immunohistochemistry for Developing an H&E-Based Mitotic Figure Detection Model

**Younghoon Kim[1], Jihye Kim[2], Gyeongsin Park[1]**
[1] Department Of Hospital Pathology, Seoul St. Mary's Hospital
[2] Center for Health Promotion and Optimal Aging, Seoul National University Hospital

**Background:**
Accurate identification of mitotic figures is essential for tumor diagnosis, grading, and prognostic assessment in various malignancies, including breast cancer, neuroendocrine tumors, gastrointestinal stromal tumors, sarcomas, and other solid tumors. However, mitotic counting on hematoxylin and eosin (H&E)-stained slides is labor-intensive and subject to substantial interobserver variability. Although phosphorylated histone H3 (PHH3) immunohistochemistry can highlight mitotic cells and improve detection, models capable of reliably detecting mitoses directly on H&E whole-slide images remain limited.

**Methods:**
A machine learning-based workflow for mitosis detection was developed using tissue microarrays (TMAs), multiplex immunohistochemistry, and digitally scanned slides of various tumors. Each core was extracted separately. Matched H&E and PHH3-stained TMA cores were digitally aligned. In QuPath, PHH3-positive mitotic cells were detected and annotated, and the corresponding annotations were transferred to matched H&E images to generate ground-truth labels. H&E image tiles containing annotated mitotic figures were extracted and used to train a deep learning model for mitotic figure detection using Neuro-T (Neurocle, South Korea). The trained model was imported into QuPath and applied to H&E TMA slides using Neuro-R.

**Results:**
In internal validation, the model achieved a precision of 91.38%, a recall of 87.10%, and an F1 score of 89.19%. In preliminary analysis of H&E TMA slides, the model detected candidate mitotic figures, including cells corresponding to PHH3-positive regions. False-positive predictions were mainly associated with apoptotic bodies, hyperchromatic nuclei, and staining artifacts, and further optimization is ongoing to improve detection performance.

**Conclusion:**
PHH3-guided multiplex immunohistochemistry provides an efficient strategy for generating objective mitosis labels on H&E images. This workflow may facilitate the development of H&E-based mitosis detection models and support more reproducible mitotic assessment in digital pathology. Further validation in larger and more diverse cohorts is warranted.

# 89 - GC-Atlas: A Large-Scale Annotated Gastric Cancer Whole Slide Image Dataset for Computational Pathology

**Wanyuan Chen[1], Yijing Jiang[2], Xinxin Ren[1], Jie Yu[1], song zhang[1], Wenran Xie[1], Yuming Bai[1], Haoda Lu[2], Xinmi Huo[2], Kokhaur Ong[2], Weimiao Yu[2], Yunzhao Chen[1]**

[1] Cancer Center, Department of Pathology, Zhejiang Provincial People's Hospital, Affiliated People's Hospital, Hangzhou Medical College

[2] Bioinformatics Institute, A*STAR

**Background / Objective:**
Gastric cancer remains a leading cause of cancer-related mortality in Asia, yet the development of AI-driven diagnostic tools is limited by the scarcity of high-quality annotated pathology datasets derived from Asian populations. Existing public datasets are predominantly Western-based, lack sufficient annotation granularity, and rarely address the spatial complexity introduced by multiple tissue sections on a single glass slide. This study aimed to construct a structured, multi-label annotated whole slide image (WSI) dataset of gastric cancer H&E-stained slides acquired at dual scanning resolutions using a proprietary semi-automatic annotation platform.

**Methods:**
A total of 2,000 H&E-stained gastric cancer WSIs from surgical resection specimens were digitized at both 40× and 20× magnification using a KF-Bio scanner. WSIs containing multiple tissue sections underwent individual fragment identification and independent annotation for each tissue section. All WSIs were imported into a proprietary web-based annotation platform incorporating a semi-automatic segmentation model that generated initial region proposals, subsequently reviewed and approved by board-certified pathologists. A structured annotation guideline was established prior to annotation, and iterative quality control procedures were performed throughout.

**Results:**
In total, 608 WSIs at 20× and 358 WSIs at 40× magnification were annotated across multiple categories, including tumour stage-related regions, histological subtypes, tumour-stroma interfaces, signet ring cell carcinoma, and necrosis. The semi-automatic platform substantially reduced average annotation time per WSI compared to fully manual workflows while maintaining pathologist-approved quality. Key annotation challenges included delineating closely apposed tissue fragments, invasion fronts, diffuse infiltrative regions, and complex tumour-stroma boundaries.

**Conclusion:**
We present a structured, multi-label gastric cancer WSI dataset derived from an Asian cohort reflecting real-world surgical pathology practice. Future work will include multi-pathologist consensus annotation, cell-level labeling, and downstream model development for tumour segmentation, subtype classification, and prognostic prediction.

# 90 - Computer-Aided Abnormality Detection of Nasopharyngeal Cancer Using Whole Slide Images

**Brian Rong Sheng Yong, Elaine Wan Ling Chan[2], Wan Siti Halimatul Munirah Wan Ahmad[2], Shew Fung Wong[1], Selvam Thavaraj[3], Lu Ping Tan[4], Geok Wee Tan[4]**

[1] Institute for Research, Development and Innovation, IMU University
[2] School of Computing and Artificial Intelligence, Faculty of Engineering and Technology, Sunway University
[3] Department of Oro-maxillofacial Surgical & Medical Sciences Faculty of Dentistry, University of Malaya
[4] Molecular Pathology Unit, Cancer Research Center, Institute for Medical Research

**Background:**
Nasopharyngeal carcinoma (NPC) is highly prevalent in Malaysia, yet frequently diagnosed at late clinical stages due to subtle early histological features. Manual screening of Whole Slide Images (WSIs) by pathologists is labor-intensive and subject to significant inter-observer variability. To address these clinical challenges, standardize pathology workflows, and assist in high-volume slide screening, we developed and validated a computer-aided digital pathology framework using deep learning to automate tumor detection in nasopharyngeal biopsies.

**Methods:**
H&E-stained formalin-fixed paraffin-embedded archival tissue slides from confirmed stages I to IV NPC cases were sourced from the archival collections at the Institute for Medical Research. Slides were digitized and tiled into patch profiles. Training utilized 30 WSIs yielding 32,642 patches (27,294 tumor; 5,348 non-tumor). Validation evaluated 19 unseen WSIs containing 28,409 patches (11,129 tumor; 17,280 non-tumor). Three modeling approaches were compared: Model 1, DenseNet-21 (DN-21); Model 2, Pyramid Vision Transformer (PVTv2); and Model 3, a hierarchical Triple-Scale DenseNet-21 Model processing 162px, 324px, and 648px patches.

**Results:**
At the patch-level, standalone PVTv2 achieved the highest overall validation accuracy of 90.09%, followed closely by DN-21 at 88.99% and the Multi-Scale Ensemble at 88.50%. PVTv2 demonstrated superior sensitivity, capturing 9,944 True Positives (recall: 89.35%) but suffered from 1,630 False Positives. Alternatively, DN-21 was highly specific with only 457 False Positives but missed substantial malignancies (2,672 False Negatives). The Triple-Scale Ensemble balanced these architectural extremes, registering 8,971 True Positives and maintaining a controlled false alarm rate of 1,110 False Positives.

**Conclusion:**
This multi-scale deep learning framework robustly demonstrates excellent clinical viability in distinguishing normal nasopharyngeal mucosa from malignant tissue. By utilizing these screening models, this platform reliably pre-analyzes large volumes of digitized archival slides, minimizing inter-observer diagnostic discrepancy, and driving the transition toward standardized workflows in clinical digital pathology.

# 93 - Resource-Agnostic Patch-Level Microsatellite Instability Prediction in Colorectal Carcinoma: A Multi-Model Foundation Benchmarking Study

**Snehvarsha Bhagat[1], Prasenjit Das[1], Rimlee Dutta[1], Devansh Lalwani[2], Swapnil Bhat[2], Maulik Shah[2]**

[1] Aiims Delhi

[2] Turocrates.ai

**Background:**

Microsatellite instability (MSI) is a critical biomarker in colorectal cancer (CRC), guiding prognosis and immunotherapy eligibility. Standard MSI testing via immunohistochemistry or PCR requires laboratory infrastructure that is frequently unavailable in resource-limited settings. Computational prediction of MSI from hematoxylin and eosin (H&E)-stained slides offers a promising alternative, but most published models depend on high-cost whole-slide imaging (WSI) scanners. This study evaluates whether a slide-level MSI prediction model trained entirely on open-source WSI data can generalize to smartphone-captured microscopy images, enabling resource-agnostic deployment.

**Methods:**

We benchmarked eight foundation model-classifier combinations for patch-level MSI classification. Extracted embeddings were paired with four classifiers: K-Nearest Neighbors, Random Forest, Logistic Regression, and Support Vector Machine, yielding eight pipelines evaluated under stratified cross-validation on the TCGA training set. The best-performing pipeline was then deployed for external validation on 250 tumor patches collected from CRC cases from a tertiary centre, using a resource-agnostic acquisition protocol to simulate clinical deployment without scanner infrastructure.

**Results:**

On external validation (N=250 tumor patches), Virchow embeddings consistently outperformed Phikon across all classifiers. The best performance was achieved by Virchow + Random Forest (accuracy 0.945, weighted precision 0.85, recall 0.83, F1 0.81), followed by Virchow + Logistic Regression (accuracy 0.82, F1 0.80). Among Phikon-based pipelines, Random Forest was strongest (accuracy 0.81, F1 0.79); KNN was weakest for both models. Across all eight pipelines, Random Forest and Logistic Regression classifiers consistently outperformed SVM and KNN, regardless of foundation model.

**Conclusion:**

Systematic benchmarking identifies Virchow + Random Forest as the optimal foundation model-classifier combination for patch-level MSI prediction. Strong external validation confirms generalizability and supports feasibility as a resource-agnostic MSI screening tool in settings where molecular testing is unavailable.

# 94 - HPLA-100: A Histopathology Performance Benchmark for LLM Assessment

**Rimlee Dutta[1], Nayantrishna Nath[2], Shraddhanjali Satapathy[3], Swapnil Bhat[4], Devansh Lalwani[4], Maulik Shah[4]**

[1] Aiims Delhi
[2] ESIC Medical College
[3] PGIMER
[4] Turocrates.ai

**Background:**
Vision-capable large language models (LLMs) are increasingly explored as diagnostic aids in histopathology, yet no standardized benchmark exists for evaluating their performance on cases relevant to low- and middle-income countries (LMICs). Existing evaluations rely on paid APIs inaccessible in resource-limited settings and lack structured assessment of safety-critical behaviors such as uncertainty management and adversarial robustness. We present HPLA-100, a curated benchmark dataset for evaluating histopathology interpretation capabilities of general-purpose LLMs under conditions reflecting real-world LMIC practice.

**Methods:**
HPLA-100 comprises 100 histopathology cases across four categories: (A) LMIC-prevalent conditions including infectious mimics and high-burden cancers; (B) resource-constrained scenarios with limited immunohistochemistry; (C) adversarial cases with impossible data or specimen mix-ups; and (D) straightforward diagnostic controls. Each case pairs a clinical vignette with a microscopic image. Five freely accessible LLMs -- ChatGPT, Gemini, Claude, Grok, and DeepSeek -- plus MedGemma- were evaluated via their free chat interfaces. Three board-certified pathologists independently scored all 600 responses on five evaluation dimensions: Correct Diagnosis, Instruction Following and Clarity, Completeness, Safety, and Uncertainty Management, using a Likert scale (-2 to +2).

**Results:**
Claude achieved the highest overall mean score (1.12), followed by Gemini (0.59), ChatGPT (0.22), MedGemma (0.13), Grok (0.05), and DeepSeek (-0.38). Claude led across all five dimensions, with particular strength in adversarial detection (1.68) and resource-constrained scenarios (1.50). On safety, Claude scored highest (0.80) while DeepSeek scored lowest (-0.69), exhibiting high failure rates and safety concerns across categories.

**Conclusion:**
HPLA-100 provides a reproducible, openly described benchmark for evaluating LLM performance in histopathology under LMIC-relevant conditions. Its four-category design stress-tests diagnostic accuracy, safety awareness, and uncertainty calibration. The benchmark reveals significant inter-model variability, underscoring the need for rigorous, context-specific evaluation before clinical deployment.

# 95 - From Cell Annotation to Molecular Subtyping: Bridging H&E Morphology and Subtypes in Invasive Breast Carcinoma

**Aida Akaeva**
Moscow Clinical Scientific Center

**Background / Objective**
Breast cancer treatment relies on molecular profiles (ER, PR, HER2) rather than morphology alone. Current IHC and FISH assessments suffer from variability due to antibody choice, staining platforms, pathologist subjectivity, and tumor heterogeneity. Existing computational pathology datasets are fragmented: they provide either nucleus annotations without clinical labels or molecular profiles without spatial context. We present a histopathology dataset pairing nucleus segmentation with slide-level clinical biomarkers, enabling the study of morphological–phenotypic relationships. The main contributions of this work are: 1) The breast cancer unique dataset pairing per-nucleus annotations with clinical labels from 80 patients; 2) Multi-task baselines jointly optimizing classification and segmentation.

**Methods**
The dataset comprises over 8,000 patches from 80 WSIs, containing 309,180 annotations across various cell types, paired with patient-level molecular subtypes. We designed multi-task deep learning baselines for joint instance segmentation and receptor status classification. To prevent the common degradation of segmentation quality during multi-task training, we "froze" the feature extractor, allowing only prediction modules to adapt. This maintains precise cell localization while ensuring accurate clinical biomarker prediction.

**Results**
We introduce the first histopathology breast cancer dataset integrating per-nucleus polygon annotations with slide-level clinical biomarkers for invasive carcinoma. This resource enables the direct study of morphology–phenotype correlations within a single cohort. Multi-task experiments demonstrate that morphological features contain predictive signals for receptor status.

**Conclusion**
Breast cancer dataset enables computational investigation of how cellular morphology relates to molecular expression, providing a foundation for holistic digital pathology workflows. Our findings suggest that H&E cell morphology contains sufficient information for receptor status, advancing the potential for AI-driven, IHC-independent receptor assessment.

# 97 - A Pathology-Specific Frechet Inception Distance for Whole Slide Image Quality Assessment Using Foundation Model Features

**Devansh Lalwani[1,2], Swapnil Bhat[2], Maulik Shah[2], Jay Mehta[3]**
[1] Seth GSMC & KEM Hospital
[2] Turocrates.ai
[3] Neuberg Oncopath

**Background:**
Quality assessment of whole slide images (WSIs) is essential for reliable computational pathology, yet standardized automated metrics remain limited. The Frechet Inception Distance (FID), widely used in generative imaging, relies on ImageNet-trained feature extractors that may fail to capture histopathology-relevant morphological characteristics. As computational pathology pipelines increasingly depend on WSI quality for downstream model performance, domain-specific quality metrics are needed. This study proposes and evaluates a pathology-adapted FID using foundation model features as an objective quality assessment tool for WSIs.

**Methods:**
Eight pathology foundation models were evaluated as feature extractors: Virchow2, Prov-GigaPath, CONCH, PLIP, UNI, Phikon, CTransPath, and RetCCL. Over 10,000 tissue-containing patches were extracted per slide and processed through each model. Feature distributions were summarized using mean and covariance matrices, and FID was computed between reference and comparison sets. Performance was benchmarked on datasets with controlled resolution degradation and JPEG compression artifacts. Concordance with pathologist visual quality assessments served as the clinical reference standard.

**Results:**
Foundation model-based FID demonstrated substantially greater sensitivity to pathology-relevant morphological differences compared with conventional ImageNet-based FID. Higher-dimensional models (Virchow2, Prov-GigaPath) detected finer-grained quality changes but produced elevated absolute FID scores requiring normalization for interpretability. Lighter-weight models (CONCH, PLIP) achieved 87% concordance with pathologist visual assessments, providing an effective balance between computational sensitivity and clinical relevance. The pathology-specific FID consistently outperformed conventional FID in distinguishing clinically meaningful quality degradation from negligible variation.

**Conclusion:**
Replacing ImageNet features with pathology foundation model embeddings substantially improves FID relevance for WSI quality assessment. Lighter models such as CONCH and PLIP offer practical, clinically concordant quality metrics suitable for integration into computational pathology quality control pipelines, with potential applications in scanner validation and laboratory accreditation.

# 98 - Cellular Sheaves for Interpretable Weakly-Supervised Pathology Localization

**Devansh Lalwani[1,2], Swapnil Bhat[2], Maulik Shah[2]**
[1] Seth Gsmc & Kem Hospital
[2] Turocrates.ai

**Background:**
Attention-based multiple instance learning (ABMIL) on foundation-model features achieves near-saturation slide-level classification on whole-slide images, yet the resulting attention maps are unreliable localizers: a model that classifies correctly without highlighting the actual lesion is difficult to trust clinically. We address this interpretability gap using cellular sheaves -- topological structures that equip graph vertices and edges with vector spaces and consistent linear maps -- providing a principled framework for detecting local disagreement on graph-structured histopathology data.

**Methods:**
We construct a spatial patch graph from UNI foundation-model features and attach a low-rank cellular sheaf whose restriction maps are produced by a hypernetwork. We introduce attention-conditional consistency, a training objective using the classifier's own attention to define which neighboring patches should agree or disagree: edges between similarly weighted patches are encouraged toward consistency, while edges across attention discontinuities are encouraged toward disagreement. The classifier and sheaf are trained jointly, allowing the shared projector to co-adapt under both objectives. This produces two complementary localization signals: the attention map (classifier focus) and the sheaf disagreement field (tissue-level consistency violations).

**Results:**
On Camelyon16 (49 tumour test slides), joint training achieved a disagreement field AUC of 0.940 and raised attention AUC from 0.717 (ABMIL-only) to 0.953. A two-stage ablation with a frozen classifier reached only 0.727 disagreement AUC, confirming that gains require joint optimization. Without retraining, the model transferred to Camelyon17 (44 tumour slides from distinct medical centres), maintaining disagreement AUC of 0.932 +/- 0.083 and attention AUC of 0.955 +/- 0.099.

**Conclusion:**
Attention-conditional consistency enables cellular sheaves to produce a diagnostically meaningful localization signal that complements and sharpens ABMIL attention. The dual-map output -- attention and sheaf disagreement firing on the same regions for different mathematical reasons -- provides pathologists with two independent explanations per prediction, supporting trustworthy interpretation of weakly-supervised whole-slide image analysis.

# 99 - Foundation Model Assisted Glomerular Detection and Feature Assessment Workflow for Lupus Nephritis Renal Biopsies

**Hao-Wen Chuang[1], Zhen-You Huang[2], Bo-Yu Chen[2], Yu-Chieh Lin[3,4]**

[1] Department of Pathology and Laboratory Medicine, Kaohsiung Veterans General Hospital
[2] Institute of Photonics Engineering, National Kaohsiung University of Science and Technology
[3] Digital Medicine Center, Translational Health Research Institute, Vilnius University
[4] Institute of Research, Development and Innovation, IMU University

**Background:**
Histopathological evaluation of Lupus Nephritis renal biopsies remains challenging due to the need for accurate identification of glomeruli and assessment of sclerosis and activity-related features for disease classification. We developed an AI-assisted workflow to support efficient glomerular evaluation in PAS-stained renal needle biopsy whole slide images (WSIs).

**Methods:**
A total of 52 PAS-stained renal needle biopsy WSIs were collected for model development. Instead of relying on conventional object detection approaches, WSIs were divided into 256×256 image patches for glomerular identification. A foundation model based framework using CHIEF CTransPath feature extraction and ConvNeXt V2 downstream classification was implemented for glomerular detection and subsequent sclerosis/activity-related feature recognition.

**Results:**
The proposed glomerular identification workflow achieved an accuracy of 92.8% and an F1-score of 92.7%, significantly improving performance compared with preliminary experiments using YOLOv11-based object detection approaches (~80% accuracy). The workflow effectively identified glomerular regions and estimated glomerular distribution within WSIs. For sclerosis and activity-related feature recognition, the same framework achieved over 80% classification accuracy in both feature categories.

**Conclusions:**
We developed a foundation model assisted workflow for LN renal biopsy assessment capable of glomerular identification and feature evaluation in PAS-stained WSIs. This workflow may improve the efficiency and reproducibility of LN pathological evaluation and provide a practical foundation for future AI-assisted LN classification systems. Future work will include larger-scale validation, integration of glomerular feature statistics with LN class prediction, and extension to H&E-stained slides.

# 100 - Validation of a Three-Way HER2 Scoring AI Algorithm in Breast Carcinoma

**GURUDUTT Gupta[1], Vinod PK[2], Anurag Mehta[1], Sunil Pasricha[1], Meenakshi Kamboj[1], Anila Sharma[1]**
[1] Rajiv Gandhi Cancer Institute & RC
[2] IIIT

**Introduction:**
Digital pathology and artificial intelligence (AI) are increasingly supporting histopathological diagnosis. We present a validation study of a three-way HER2 scoring algorithm developed through an academic collaboration between RGCIRC and IIIT Hyderabad, using whole-slide imaging (WSI).

**Methods:**
The HER2 classifier, developed by IIIT Hyderabad, was deployed as a web-based application. A total of 200 consecutive retrospective HER2 IHC breast carcinoma cases were analyzed over four months. WSIs were uploaded and algorithm outputs recorded. Ground truth was established from routine reporting by two onco-pathologists. Discordant cases were jointly reviewed, and consensus scores were assigned where applicable.

**Results:**
All 197 of 200 cases were successfully processed. The AI algorithm demonstrated moderate-to-substantial agreement with pathologist scoring, with ICC values in the moderate-to-good range. Agreement improved measurably after HER2 Algorithm Validation consensus review, driven primarily by pathologist reclassification of borderline HER2 0 cases to HER2 Low — a category the algorithm had already flagged. The most challenging boundary remains HER2 0 versus HER2 Low, which is known to be the most subjective distinction in HER2 scoring. Discussion: This study highlights the importance of site-specific validation before deployment of AI tools in clinical practice. The results demonstrate strong agreement with expert scoring, supporting the algorithm's utility as an assistive tool. While three-way classification is adequate for trastuzumab eligibility, emerging therapies require more granular classification.

**Conclusion:**
The validated AI algorithm shows promising accuracy and clinical applicability in HER2 scoring. Continued multi-center validation and expansion to four-way classification are next goal. Keywords: HER2, Digital Pathology, Artificial Intelligence, Breast Cancer, Validation Study

# 102 - When Digital Pathology Thinks Like a Pathologist: An Aiforia Experience in Breast Biomarker Assessment in a High-Volume Diagnostic Laboratory

**Nidhi Garg, Priyanka Kembhavi, Brijesh Kumar, Swathi Vadlamudi, Kush Raut, Kunal Sharma**
Centre of Excellence – Histopathology, Agilus Diagnostics.

**Background**
Accurate assessment of breast cancer biomarkers including estrogen receptor (ER), progesterone receptor (PR), HER2, and Ki67 is essential for therapeutic decision-making. However, manual interpretation is associated with inter-observer variability, particularly in borderline and heterogeneous cases. This pilot study evaluated the utility of the Aiforia AI platform for digital quantification of breast cancer biomarkers in a routine diagnostic setting.

**Methods**
A pilot cohort of 50 breast carcinoma cases, including core biopsies and resection specimens, was evaluated. Whole-slide images were analyzed using the Aiforia AI platform integrated into a PathPresenter-based digital pathology workflow. Three pathologists independently assessed biomarker expression manually, and the results were compared with AI-assisted quantification. Concordance, reproducibility, turnaround time (TAT), and workflow efficiency were assessed.

**Results**
Manual reporting demonstrated a 4% discordance rate for HER2 and a 6% discordance rate for Ki67. The AI-assisted workflow resolved these discrepancies, significantly reducing intra and inter-observer variability while demonstrating high reproducibility. Automated hotspot detection and tumour segmentation enabled objective Ki67 quantification, particularly in heterogeneous tumours. For ER and PR, manual Allred scores demonstrated 100% concordance, with only minor clinically insignificant variability in percentage estimation of positive cells. Operationally, the automated AI workflow reduced turnaround times by 54%, with all cases completed within the optimised TAT. The digital pipeline also achieved 100% compliance with Proficiency Testing (PT) and External Quality Assessment Scheme (EQAS) metrics.

**Conclusion**
Integrating AI into digital pathology may improve standardisation of challenging breast biomarker quantification while reducing clinically relevant inter-observer discordance. By reducing turnaround times by more than 50%, this automated workflow may improve resource utilization and support high-volume laboratories through standardised and reproducible clinical quality assessment while retaining pathologist oversight.

# 104 - Real-World Insights in Designing an End-to-End Deep Learning Pipeline for Hepatic Steatosis Quantification

**Nagalakshmi Jegannathan, Wei Qiang Leow**
Singapore General Hospital

**Introduction:**
Metabolic dysfunction-associated steatotic liver disease (MASLD), with an estimated prevalence of 30%, is a significant and escalating global health concern. Hepatic steatosis, a hallmark of MASLD, is currently graded semi-quantitatively by pathologists; however, this method is limited by inter-observer variability. To address this limitation, we proposed to develop an end to end pipeline to quantify hepatic steatosis. Methodology A total of 38 archived liver core biopsies from MASLD patients (2005–2017) were processed using routine clinical protocols (4 µm H&E sections) and independently assessed by three expert pathologists for fat percentage and NASH CRN grading. The dataset included 20 WSIs for training and validation (12,154 annotated ROIs) and 18 WSIs for testing. WSIs were converted to BigTIFF, tiled into 1024×1024 patches, and annotated for seven liver structures, with separate U-Net models trained for each. Patch-level predictions were stitched to compute total structure areas, and fat percentage (fat area/parenchyma area) was mapped to NASH CRN steatosis grades.

**Results**
SteatoStat achieved a weighted kappa of 0.669 versus expert pathologists, with a mean fat percentage difference of −0.01% (95% CI: −29.3% to +29.3%). Compared to an established machine learning algorithm, it showed higher agreement (weighted kappa 0.850) and minimal bias (mean difference 1.85%, 95% CI: −20.3% to +24.0%).

**Conclusion**
Previous studies have applied AI for steatosis quantification using different approaches. Goh et al. used SHG and TPEF microscopy on unstained slides, which requires specialised imaging; in contrast, we leverage standard H&E-stained WSIs from clinical practice. Forlano et al. calculated fat percentage relative to total tissue area, potentially underestimating steatosis, whereas our pipeline excludes non-parenchymal structures for improved accuracy. Overall, we developed an end-to-end deep learning pipeline that shows strong agreement with expert pathologists and an established machine-learning method, while addressing real-world variability such as multiple WSI formats and multi-structure segmentation.

# 105 - DURIAN (Dual-Stain Renal Image Analysis Network): a phenotype-aware virtual staining framework for application in Diabetic Kidney Disease

**Sahana Abdul Rahim[1], Charlene Zhi Lin Ong[1], Kelvin Kian Long Chong[1], Traci Ann Czyzyk[2], Chih-Liang Chin[2], Shubing Wang[3], Richard Baumgartner[3], Asad Abu Bakar Ali[1], Desiree Abdurrachim[1]**

[1] Quantitative Biosciences, MSD

[2] Cardiometabolic Diseases, Merck & Co., Inc.

[3] Biostatistics and Research Decision Sciences, Merck & Co., Inc.

**Background:**

Diabetic Kidney Disease (DKD) is a multifactorial and heterogenous condition that is an increasing global health concern, affecting over 40% of diabetic patients. In preclinical research, the disease understanding typically requires histopathological assessments using Hematoxylin and Eosin (H&E) staining as well as multiple immunohistochemistry (IHC) special stains to directly target biomarkers of interest. However, acquisition of such stains can be time consuming and laborious, motivating the need for virtual staining of H&E. Here, we aim to develop AI virtual staining methods to generate Collagen-IV (Col-IV) IHC images from H&E.

**Methods:**

We present DURIAN, a phenotype-conditioned generative adversarial framework with addition of two key innovations: (i) a modified protein-aware loss from empirical data, and (ii) a novel mechanism of incorporating the disease label into the generator, to make the AI model generalizable for different phenotypes which otherwise may require separate AI model training. The model was trained on diabetic and non-diabetic datasets using 3892 and 2528 patches for training, and 5522 and 3648 patches for testing respectively. We assess the model performance on three levels: image quality, biological structural fidelity, and ability to retain Col-IV stains between real and virtual images.

**Results:**

Across diabetic and non-diabetic datasets, DURIAN produces images of high quality, with strong tubule and glomerular fidelity (diabetic/non-diabetic: LPIPS=0.298/0.303, tubules IoU=0.89/0.90, glomerulus IoU=0.91/0.92). Compared to the model without the two additions, DURIAN generates the virtual Col-IV intensity with lower error for diabetic dataset (1.8-fold) and non-diabetic dataset (8-fold) with a higher correlation (1.2-fold) to the true intensity for non-diabetic dataset. DURIAN outperforms phenotype-specific models, reducing Col-IV intensity error by 2-fold for diabetic and non-diabetic datasets, which suggest the capability of DURIAN in generalizing across animal model phenotypes.

**Conclusion:**

DURIAN is a promising approach for H&E-to-IHC staining, contributing to time and cost-efficient tissue analysis approach for DKD research.

# 106 - AI detection and quantification of retinal vasculature using the STZ animal model

**Denise Ng[1], Stephanie Cicalese[2], Asad Abu Bakar Ali[1], Weiwei Luo[1], Xiaolan Shen[2], Desiree Abdurrachim[1]**
[1] MSD, Singapore
[2] Merck & Co., Inc., Rahway, NJ, USA

**Background:**
Diabetic retinopathy (DR) is a common microvascular complication of diabetes and a leading cause of irreversible vision loss. It is marked by vascular abnormalities such as microaneurysms, hemorrhages, capillary occlusion, macular edema, and neovascularization. In this study, we investigate vascular pathology in a preclinical rodent model using an AI-based method that removes the need for manual analysis, corrects histology artifacts, and efficiently quantifies vessel density and branching.

**Methods:**
Diabetic retinopathy was induced in Brown Norway rats via intraperitoneal streptozotocin (STZ) injection. Eyes were collected 1(n=4 Vehicle, n=9 STZ) and 4 months (n=8 Vehicle, n=11 STZ) post-injection and dissected after incubation in 4% PFA. Retinal blood vessels were then visualized with Isolectin B4 staining and imaged at 40x magnification on an Axioscanner. A two-stage deep learning model was developed to quantify vascularization from the flatmount images. Both stages use a U-Net architecture with a ResNet-50 backbone for semantic segmentation. The first stage segments tissue and removes artifacts from downsampled images (N=51 training, N=13 validation), while the second stage detects vessels within these areas using full-resolution 1024x1024 patches (N=361 training, N=91 validation). Skeletonized vessel masks were used to calculate branch point density and reticulation density relative to tissue area.

**Results:**
The AI model achieved high performance with IoU of 0.911 (stage 1) and 0.689 (stage 2). At 1 month, branch point (P=0.621) and reticulation (P=0.673) densities did not differ between STZ and naïve animals. However, 4-month STZ animals had higher branch point density (8% vs. 4-month naïve, P=0.045; 12% vs. 1-month STZ, P=0.002) and higher reticulation density (5% vs. 4-month naïve, P=0.0098; 5% vs. 1-month STZ, P=0.0095), indicating successful neovascularization.

**Conclusion:**
We developed a robust deep learning model for retinal vascular analysis, enabling faster preclinical characterization and pharmacodynamic assessment of vascular pathology in MSD's ophthalmology portfolio.

## 107 - Comprehensive clinical evaluations of 38 digital pathology slide scanners

**Adler Ju, Kenny Lou, Ronald CK Chan**
Chinese University Of Hong Kong

**Background:**
Digital pathology slide scanners are an essential tool in digital pathology transformation but it can be difficult to evaluate and compare these complex devices. This study aims to independently evaluate all major scanner devices in an in-depth clinical focused method.

**Methods:**
38 digital pathology slide scanners were evaluated with over 10 clinical-grade parameters in how well they produce digital images from glass slides for diagnostic assessment.

**Results:**
A wide range of results in colour accuracy, scanning speed, focus and tissue detection success, subjective resolution graded by pathologists, WSI file workability and many other metrics are measured and shared.

**Conclusions:**
A wide range of results are displayed across many scanners. No one scanner excels in all categories but it is helpful to compare and understand the importance of each metric and the needs of each laboratory.

## 110 - Interactive Auto-Segmentation on Whole Slide Histopathology Images Using SAM3: Streamlining Digital Pathology Workflows

**Jun Aun Tan, Yijing Jiang, Haoda Lu, Kok Haur Ong, Weimiao Yu**
A*STAR Bioinformatics Institute

**Background/Objective:**
Precise annotation of cellular structures, complex glands, and tissue constituent boundaries in digital pathology is traditionally a labor-intensive process heavily reliant on tedious freehand drawing by pathologists, leading to annotation fatigue. Our objective was to present a modern digital pathology solution that eliminates the need for manual free-hand drawing by leveraging the Segment Anything Model 3 (SAM3) to empower users with an interactive, highly accurate auto-segmentation workflow directly within a web-based WSI viewer.

**Methods:**
We engineered a seamless pipeline integrating a fine-tuned SAM3 architecture into our cloud-ready digital pathology platform (AiHistoCloud). The UI/UX retains traditional free-hand annotation tools while introducing an optional SAM3-assisted segmentation mode, enabling users to define regions of interest (ROIs) by simply drawing a bounding box or highlighting target areas directly on the histopathology image. The system performs real-time server-side inference using the SAM3 foundation model on dynamically retrieved high-resolution image patches and converts the resulting masks back into vector polygons mapped onto the viewer in real-time.

**Results:**
The implementation successfully replaced manual outlining with a rapid prompt-to-mask workflow. SAM3's highly generalized zero-shot capabilities accurately segmented a diverse array of target structures ranging from isolated individual cells/nuclei to complex, tortuous glandular formations and broader tissue regions. By conducting inference directly on original-resolution WSIs, the boundary fidelity of the segmentations remained diagnostically relevant. Preliminary workflow analysis indicated a substantial reduction in annotation time per slide compared to traditional drawing modalities.

**Conclusion:**
Integrating SAM3 via a prompt-based auto-segmentation pipeline directly on large, uncompressed histopathology images represents a critical advancement for digital pathology. By shifting the user paradigm from granular free-hand drawing to rapid ROI highlighting, this workflow alleviates the annotation burden on pathologists, accelerates the generation of high-quality datasets for AI research, and establishes a highly scalable and workflow-efficient foundation for next-generation digital pathology solutions.

# 113 - Elucidation of Retinal Spatial Compartmentalization for Identification of Neovascular Events in Pre-clinical Models

**Kelvin Chong[1], Alexander Loktev[2], Xiaomeng Wang[4,5,6], Alex M Tamburino[3], Asad Abu Bakar Ali[1], Ser Mien Chia[1]**

[1] Quantitative Biosciences, MSD International GmbH
[2] Discovery Biology Merck & Co., Inc
[3] Data, AI, and Genomic Sciences, Merck & Co., Inc
[4] Centre for Vision Research, Duke-NUS Medical School
[5] Singapore Eye Research Institute
[6] Institute of Molecular and Cell Biology, A*STAR

**Background/Objective:**
Abnormal vascular growth in the retina (retinal neovascularization) has been linked to blinding disease of the eye such as diabetic retinopathy, diabetic macular edema and choroidal neovascularization. Despite the widespread use of pre-clinical models, spatial organization of retinal neovascularization and its associated mechanisms have not been studied. Our aim is to map cellular dysfunctions and interactions in the retinal layers with the aim of identifying targets to prevent disease onset and progression.

**Methods:**
We utilize a MERFISH-based spatial transcriptomics platform to uncover the spatial organization and resolve the cellular identities in the retina. We leverage on the mouse oxygen-induced retinopathy model which resulted in formation of avascular zones and neovascular tufts, which are hallmarks of retinal neovascularization. We designed a panel of 500 genes targeting the 13 major cell types in the retina and probed for hypoxia-driven events underlying disease development and progression (E.g hypoxia, immune function, and metabolic genes). By integrating transcriptomics and histopathology, we performed combinatorial cellular segmentation and identification, molecular signature imputation, and receptor-ligand interactions to delineate neovascular-contributing mechanisms.

**Results:**
We demonstrated the detection of major retinal cell-types including cell-types with limited markers (E.g RPE cells, <3 markers). We were able to detect inter-endothelial interactions between VEGF-related and independent pathways, which were previously confounded due to lack of spatial information in scRNA studies. To the best of our knowledge, we report the first spatial profiling of hypoxia-controlled genes in retinas from neonatal, which will help us understand the cellular mechanisms involved in ischemic retinal diseases.

**Conclusion:**
Our ability to spatially profile the retina allow us to resolve both the cellular identities and its respective functional niches within the diseased retinas. Taken together, our goal is to create a comprehensive map of these hypoxic events to probe underlying disease mechanisms and monitor intervention strategies.

# 115 - AI-Assisted Diagnostic Classification of BCR::ABL-Negative Myeloproliferative Neoplasms Using a Pathology FoundationModel with Attention-Based Multiple Instance Learning

**Liqun Yang, Ying Liang, Zuhao Huang, Ming Ding, Chunbo Huang**
Guangzhou Huayin Medical Laboratory Center Ltd.

**Background**
BCR::ABL-negative myeloproliferative neoplasms (MPN) primarily include essential thrombocythemia (ET), polycythemia vera (PV), prefibrotic primary myelofibrosis (Pre-PMF), fibrotic primary myelofibrosis (Fibrotic PMF), and MPN, not otherwise specified (MPN, NOS). Accurate histopathological classification remains challenging owing to overlapping morphological features and substantial inter-observer variability.

**Methods**
A total of 2,049 bone marrow whole slide images (WSIs) were used for binary classification (MPN: n=1421; non-MPN: n=628), and a subset of 1,320 MPN WSIs were further used for five-class subtyping (ET: n=362, PV: n=284, Pre-PMF: n=251, Fibrotic PMF: n=410, MPN, NOS: n=13). All slides were annotated by senior hematopathologists and divided into training/test sets (80/20 split). The test set for five-class subtyping comprised 264 WSIs: ET (n=72), PV (n=57), Pre-PMF (n=50), Fibrotic PMF (n=82), and MPN, NOS (n=3). Feature extraction was performed using PanoPath (developed by Guangzhou Huayin Medical Laboratory Center Ltd.), a multimodal pathology foundation model pre-trained on >3M WSIs. An attention-based multiple instance learning (MIL) framework was applied for slide-level feature aggregation.

**Results**
Binary classification achieved 81.7% accuracy (weighted F1=0.81). For MPN, sensitivity was 93% and precision 83%; for non-MPN, precision was 79% and recall 56% (Figure 1, Table 1). In five-class subtyping, overall accuracy reached 67.4% (weighted F1=0.67; macro F1=0.52) (Figure 1, Table 1). Fibrotic PMF achieved the highest F1 (0.87), followed by ET (0.64) and PV (0.62). Pre-PMF showed moderate performance (F1: 0.48), reflecting morphological overlap with ET and PV. MPN, NOS received 0 predictions due to insufficient training samples (n=10); exclusion from macro F1 calculation yields macro F1=0.65.

**Conclusion**
The PanoPath-based MIL framework demonstrates robust performance for both MPN screening and subtype classification, particularly achieving high diagnostic accuracy for Fibrotic PMF (F1=0.87). Remaining challenges in differentiating morphologically similar subtypes (e.g., Pre-PMF vs. PV) support future work with larger datasets and multimodal fusion of histologic, clinical, and molecular data.

# 116 - Japan-Based Multimodal Medical Image Data Platform for AI R&D and Clinical Research

**Changhee Han**[1,2,3]
[1] Callisto Inc.
[2] Osaka University
[3] Nagasaki University

**Background / Objective:**
Artificial Intelligence (AI)-based medical devices and AI-enabled drug discovery increasingly depend on large, high-quality medical image datasets. However, Asian datasets remain underrepresented, and Whole-Slide Images (WSIs) linked to clinical and molecular variables are difficult to prepare for Research and Development (R&D), because acquisition, anonymization, standardization, annotation, and quality control require integrated medical and AI expertise. We developed a Japan-based radiology and pathology data platform and expanded it into multimodal dataset infrastructure for medical AI and drug-discovery R&D.

**Methods:**
We implemented an end-to-end workflow covering medical-institution onboarding, anonymization, AI-assisted case/series/slide selection, image and metadata standardization, expert annotation, quality control, and dataset management. The platform covers radiography, computed tomography, magnetic resonance imaging, positron emission tomography, mammography, ultrasound, and WSIs, together with radiology/pathology reports, clinical variables, and molecular data when available. Internal tools support image and clinical-text de-identification, preprocessing, annotation management, and series/slide-level inventory linked to a data portal.

**Results:**
Over the past two years, the platform has evolved from an initial data marketplace concept into operational infrastructure for AI R&D and AI-enabled drug discovery. Data-supply relationships have expanded to major Japanese hospitals and clinics with large radiology and pathology archives. Collaborations with teleradiology, pathology AI, and imaging-CRO companies have strengthened data acquisition, processing, and distribution. Curated datasets have been provided to approximately 20 stakeholders, including medical AI companies, medical device manufacturers, pharmaceutical companies, and academic institutions in Japan, the United States, Europe, and Korea. Current priority datasets include lung, breast, prostate, colorectal, gastric, brain, liver, and normal-control categories across radiology and pathology.

**Conclusion:**
This platform provides an integrated Japan-based pathway for transforming real-world medical imaging resources into quality-controlled multimodal datasets. By supplying underrepresented Asian radiology and pathology data with clinical and molecular context, it may accelerate digital pathology, medical AI, and AI-enabled drug-discovery R&D.

# 119 - A Vision-Language Pathology Foundation Model for Colorectal Cancer Diagnosis and Prediction

**Li Liang[1,2,4], Zhengyu Zhang[1,2], Zhenfeng Zhuang[3], Liansheng Wang[3]**
[1] Department of Pathology, Nanfang Hospital, Southern Medical University
[2] Guangdong Provincial Key Laboratory of Molecular Tumor Pathology
[3] Department of Computer Science at School of Informatics, Xiamen University
[4] Jinfeng Laboratory

**Background:**
Accurate pathological diagnosis of colorectal cancer (CRC) is vital for treatment planning and prognosis. Yet routine pathology is under pressure from increasing sample volumes and a global shortage of pathologists. Existing CRC task-specific AI models may not support comprehensive diagnostic reasoning and have rarely been prospectively validated in real-world workflows. We aimed to develop and clinically validate a CRC foundation model that can address broad pathology tasks.

**Methods:**
We present CRChat, a vision–language foundation model for CRC pathology. A CRC dataset of 58,092 WSIs from 17,500 patients across 12 cohorts was assembled, covering both surgical resection and biopsy specimens. Prospective validation was conducted using ClinicalTrials.gov-registered data (NCT07239063). Built upon the LLava framework and unitialized from Qwen 2.5, CRChat integrates a large vision transformer (ViT-L/16) . CRChat was trained for 17 diagnostic/predictive categories, including histological subtyping, tumor regression grade (TRG), mucosal biopsy diagnosis, key gene mutation prediction, and prognosis. To align with clinical practice, CRChat also supports visual question answering (VQA) and end-to-end standardized pathology report generation.

**Results:**
Across evaluations, CRChat ranked first overall against five pathology foundation baselines across all tasks. In addition, performance was supported by prospective, workflow-based validation, demonstrating robustness under real clinical conditions. For the histological grading of adenocarcinoma NOS, CRChat achieved an average macro-AUC of 0.872 in five internal validation and reached a peak macro-AUC of 0.840 in five external validations. In the two survival analysis datasets, CRChat achieved results with C-index of 0.652 and 0.640 respectively. Among all prospective diagnostic tasks, CRChat achieved the highest AUC of 0.846 in the histological typing classification. For predictive tasks, the model reached its best performance in MMR status prediction, with an AUC of 0.871.

**Conclusion:**
CRChat shows that disease-specific vision–language foundation models can unify diverse CRC pathology tasks and enable clinically meaningful, scalable AI translation in pathology.

# 121 - Deep Learning-Based Virtual Immunohistochemical Staining of Mismatch Repair Proteins in Colorectal Cancer

**Zhijian Cen[1], Linbo Liao[2], Zhengyu Zhang[1], Liansheng Wang[2], Li Liang[1], Yaping Ye[1]**
[1] Southern Medical University
[2] Xiamen University

**Background:**
Microsatellite instability (MSI) is a key predictor of immunotherapy response in colorectal cancer (CRC). However, current immunohistochemistry (IHC) for mismatch repair (MMR) proteins is limited by operational complexity and sample quality dependence.

**Objective:**
To develop a generative adversarial network (GAN)-based virtual staining model that generates MMR protein IHC images from hematoxylin and eosin (H&E) slides of CRC.

**Methods:**
Consecutive H&E and MMR IHC slides from 40 CRC cases (200 slides) were collected, registered, and used to train multiple GAN models. Six pathologists evaluated the virtual images via a visual Turing test. The optimal model was validated against genetic testing and spatial transcriptomic data.

**Results:**
1.A total of 40 CRC cases with consecutive H&E and MMR protein IHC slides were collected, comprising 200 histopathological slides. Ten feature points were annotated per slide, totaling 2,000 points for model registration. 2.According to six pathologists' evaluation, the deformation-registered CUT model produced virtual IHC images that were judged as real with an average rate of 88.3% and met clinical usage standards at an average rate of 93.6%, making it the most suitable model for consecutive slide virtual staining. 3.For MLH1 virtual staining consistency analysis, using a tumor cell positivity threshold of 10%, the Kappa value was 0.958, the concordance rate was 93.75%, and $P < 0.001$. Moreover, the virtual staining model showed a degree of colocalization with MMR gene expression in multi-platform spatial transcriptomic data.

**Conclusion:**
The deformation-registered CUT model generates reliable virtual MMR IHC staining from H&E images, enabling accurate MSI status assessment in CRC. This approach reduces time and material costs and offers a new method for virtual staining on consecutive pathological slides. Keywords: Microsatellite instability; Colorectal cancer; Artificial intelligence; Virtual staining

# 124 - Annotation-Free Prediction of 5-Year Renal Function Decline in IgA Nephropathy: A Multiscale Teacher-Student Distillation Framework

**Sungman Hong[1], Kyung A Kim[2], Sang-Jung Ahn[3], Hyo Sup Shim[2]**

[1] Department of Biomedical Informatics, College of Medicine, Korea University

[2] Department of Pathology, Yonsei University College of Medicine

[3] Department of Pathology, Korea University Anam Hospital

**Background/Objective:**

IgA nephropathy (IgAN) is the most common primary glomerulonephritis, with 20-40% of patients progressing to kidney failure within two decades. While validated risk tools have shown utility for IgAN prognosis, extensions incorporating the Oxford MEST-C classification further integrate histopathologic information through manual lesion scoring and segmentation-based pipelines. Severe class imbalance (11:1, negative-to-positive) in progression cohorts further destabilises model training. Unlike Oxford MEST-C segmentation-based risk stratification, we developed an annotation-free, end-to-end framework for predicting 5-year renal function decline directly from PAS-stained WSIs.

**Methods:**

A two-stage teacher–student framework was trained on 764 PAS-stained WSIs from 940 IgAN patients. A multiscale ensemble teacher aggregated patch embeddings from CONCH v1.5 and TITAN pathology foundation models across three magnifications (10×, 20×, 40×) using attention-based multiple instance learning (ABMIL) with negative downsampling (1:4), producing ensemble-averaged soft logits. These were distilled into a compact fusion student model trained on the full imbalanced distribution, stabilising decision boundaries. Segmentation models (DeepLabV3+) derived lesion area-ratio features for comparative structural analysis.

**Results:**

Teacher–student distillation improved the composite Score from 0.668 to 0.764 on Test Phase 1 (AUC: 0.883, F1: 0.684), a 14.4% relative gain, with Test Phase 2 yielding Score 0.595, reflecting cohort-level domain shift. This framework ranked first in the 2025 MAIC Challenge (97 teams, 8 finalists). Segmentation-derived lesion area ratios (IFTA, global sclerosis, interstitial inflammation) showed modest correlations with outcome (Spearman $\rho \leq 0.252$) and provided ,by themselves, no improvement over the WSI-only baseline. For model validation, attention maps concentrated on IFTA and interstitial inflammation, confirming spatial alignment with established prognostic lesions.

**Conclusion:**

An annotation-free teacher-student distillation framework enables IgAN risk prediction from PAS WSIs without pixel-level annotation. Knowledge distillation effectively stabilises learning under severe class imbalance, and attention analysis confirms alignment with established histopathological lesion patterns. Improving external generalisability through domain adaptation represents an important future direction.

# 126 - Quantitative Assessment of Nucleolar Prominence in Clear Cell Renal Cell Carcinoma Using Computer Vision and Digital Pathology

**Elena Ivanova, Alexey Fayzullin, Peter Timashev**
Sechenov University

**Background / Objective**
The WHO/ISUP grading system for clear cell renal cell carcinoma (ccRCC) is primarily based on the assessment of nucleolar prominence. However, visual evaluation of nucleoli is subjective and prone to interobserver variability. This study aimed to develop and validate a computer vision model for automated detection of tumor cells with prominent nucleoli in whole-slide images (WSI) of ccRCC and to evaluate the prognostic significance of quantitative nucleolar metrics.

**Methods**
A total of 144 hematoxylin-and-eosin–stained ccRCC whole-slide images from three independent cohorts were included. Tumor regions were annotated by expert pathologists. Nuclear segmentation was performed using the StarDist algorithm, followed by cell classification with a ResNet18-based model. Tumor cells were categorized according to nucleolar morphology, and quantitative parameters describing the density and proportion of nucleolus-positive cells were calculated. Model performance was assessed using receiver operating characteristic (ROC) analysis. Associations between cellular composition, tumor grade, and patient survival were evaluated using statistical and survival analyses.

**Results**
The developed model successfully identified tumor cells with visible nucleoli and distinguished high-grade from low-grade tumors with an AUC of 0.79. Quantitative analysis demonstrated a significant increase in the proportion of nucleolus-positive cells with increasing WHO/ISUP grade. Considerable intratumoral and intertumoral heterogeneity was observed, particularly among grade 3 tumors. Survival analysis revealed that tumors containing more than approximately 10% nucleolus-positive cells were associated with significantly poorer overall survival. Furthermore, four distinct histological patterns based on the relative abundance of nucleolus-positive and other tumor cells were identified, each demonstrating different prognostic outcomes.

**Conclusion**
Computer vision–based quantitative assessment of nucleolar prominence provides an objective and reproducible approach to ccRCC grading. Beyond conventional WHO/ISUP classification, quantitative cellular composition analysis offers additional prognostic information and may support future refinement of renal tumor grading systems and personalized patient management.

# 128 - Multimodal Artificial Intelligence Predicts PIK3CA Mutation in Breast Cancer from Digital Pathology and Clinical Data: A Multicenter Study

**Jiaxian Miao[1], Qi Liu[2], Jianing Zhao[1], Shishun Fan[1], Shenwen Wang[2], Feng Ye[3], Si Wu[1], Jinze Li[1], Huirui Zhang[1], Meng Zhang[4], Hong Bu[3], Xiao Han[5], Lianghong Teng[4], Yueping Liu[1]**

[1] Department Of Pathology, Hebei Medical University Fourth Hospital
[2] School of Information Engineering, Hebei GEO University
[3] Laboratory of Pathology, West China Hospital, Sichuan University
[4] Department of Pathology, Xuanwu Hospital, Capital Medical University
[5] College of Biomedical Engineering, Sichuan University

**Objective:**
Accurate detection of PIK3CA mutations is essential for guiding PI3K-targeted therapies in breast cancer, yet sequencing is not universally accessible, and single-modality prediction models have limited performance. This study developed a multimodal deep learning framework integrating whole-slide imaging (WSI) and structured clinical data to improve mutation prediction.

**Methods:**
A total of 1,047 patients from TCGA and 166 patients from 3 external centers were included. The histopathology model used a transformer-based pretrained encoder (H-optimus-0) and a clustering-constrained attention multiple instance learning (CLAM-SB MIL) classifier to generate WSI-level representations. The clinical model incorporated engineered clinical variables and an extreme gradient boosting (XGBoost) model. A decision-level late fusion strategy (Multimodal PIK3CA Model, MPM) combined probabilistic outputs from both branches. Performance was evaluated with the area under the curve (AUC) and secondary metrics. Interpretability was assessed via attention heatmaps and shapley additive explanations (SHAP) analysis.

**Results:**
MPM outperformed single-modality models. It achieved an AUC of 0.745 on TCGA and maintained stable performance across external cohorts (0.695, 0.690, and 0.680). SHAP analysis identified molecular subtype as the most influential clinical feature, whereas attention maps highlighted mutation-associated morphological regions.

**Conclusions:**
The developed multimodal framework effectively integrates complementary morphological and clinical information, and provides a robust and generalizable method for predicting PIK3CA mutation status. Strong multicenter adaptability and biological interpretability support its potential use as a clinical decision-support tool and an accessible alternative to molecular testing.

# 131 - BCNB-bench: A Pathological Multiple Instance Learning Benchmark for Eight-Task Prediction in Breast Pathology

**Masataka Kawai[1], Hirohiko Miyake[1], Genichiro Ishii[1], Shumpei Ishikawa[2]**
[1] Department of Pathology and Clinical Laboratories, National Cancer Center Japan Hospital East
[2] Exploratory Oncology Research & Clinical Trial Center, National Cancer Center Japan

**Background:**
Prediction from hematoxylin and eosin–stained whole-slide images has gained attention with the emergence of pathology foundation models and multiple instance learning (MIL). However, reliable public benchmarks for evaluating MIL pipelines remain limited. We developed BCNB-bench, a MIL benchmark based on the public Breast Core Needle Biopsy dataset of 1,058 cases.

**Methods:**
We defined eight binary prediction tasks: estrogen receptor positivity, HER2 positivity, Ki-67 index >30%, age >55 years, lymph node metastasis, histological grade 3, invasive ductal carcinoma versus others, and tumor size >2 cm. Missing histological grade labels for 132 cases were manually complemented. We performed four-fold cross-validation across three random splits. Tumor-patch and whole-tissue tokenization strategies were compared. UNIv1, CONCH, H-optimus-1, and Virchow2 were evaluated as encoders. Max pooling, average pooling, and two ABMIL variants were assessed as MIL aggregators. Linear models were fitted to estimate the contribution of tokenization, encoder, and MIL aggregator to AUROC for each task.

**Results:**
Optimized MIL pipeline with variable-length token batching and task/hyperparameter multiplexing achieved ~2000× throughput improvement over classical CLAM-style training. H-optimus-1 on tumor patches aggregated with average pooling showed the strongest overall performance. Tumor size was excluded for summarization as it was poorly predictable. ANOVA revealed task-dependent contributions of pipeline components: tokenization had the largest effect on age prediction ($\eta p^2$=0.680), whereas encoder choice dominated lymph node metastasis prediction ($\eta p^2$=0.498). MIL aggregation generally had a modest contribution ($\eta p^2$<0.05). We defined a composite metric of contributing tasks for encoders and MIL aggregators. The best encoder H-optimus-1 achieved AUROC of 0.8179 across seven tasks, while the best MIL aggregator max pooling achieved 0.8500 across three tasks.

**Conclusion:**
BCNB-bench provides a large public MIL for pathology. Our benchmark design shows that the dominant performance driver varies by task, highlighting the importance of separating tokenization, encoder, and MIL aggregator effects for fair pipeline evaluation.

# 134 - A Code-Free Recursive Workflow for AI-Assisted Digital Pathology

**WenYih Liang[1], Hsiang-Sheng Wang[2]**
[1] Taipei Veterans General Hospital
[2] Linkou Chang Gung Memorial Hospital

**Background / Objective**
Integrating artificial intelligence (AI) into routine pathology is often hindered by the requirement for extensive technical expertise, large-scale curated datasets, and high-performance programming support. Pathologists frequently face significant delays between identifying a clinical diagnostic need and deploying a usable AI tool. The objective of this project is to introduce EasyPath, a user-friendly, code-free digital pathology platform designed to empower pathologists to independently build, train, and refine AI models through a continuous, recursive annotation–training–correction workflow.

**Methods**
EasyPath is a browser-based, scalable platform built using Python and integrated with C/C++ backend libraries to support diverse whole-slide image (WSI) formats. The platform enables pathologists to annotate regions (polygons/points), objects (tags), or whole slide images (tags) and perform code-free model training with adjustable hyperparameters. A defining feature is its recursive loop, where false-positive and false-negative predictions identified during routine sign-out are re-annotated and incorporated into subsequent training generations to iteratively enhance model performance.

**Results**
The platform demonstrated robust performance, successfully supporting 13 WSI formats and 19 representative AI models across various diagnostic tasks, including biomarker quantification, infectious disease screening, and tumor segmentation. Annotation-rendering benchmarks confirmed that WebGL-based processing allows for the display of up to 2,000,000 points with minimal latency. Furthermore, a coarse-to-fine inference workflow optimizes computational efficiency by screening at low magnification before applying high-resolution models.

**Conclusion**
EasyPath effectively bridges the gap between AI research and clinical practice by removing the coding barrier and enabling pathologist-driven, continuous model refinement. By facilitating rapid development, cross-institutional model weight transfer, and real-world feedback, EasyPath provides a sustainable infrastructure for integrating AI into routine, high-volume pathology practice.

# 137 - Deep Learning-Based Classification of Peripheral Blood Cells Using EfficientNet-B0: Emphasis on Granulocytic Maturation Stages

**Naveen K Devadas, Parikshit Sanyal, Venkatesan Somasundaram**
Armed Forces Medical College, Pune

**Background:**
Peripheral blood smear examination remains a cornerstone of hematological diagnosis. Although artificial intelligence (AI)-based blood cell classification has demonstrated promising results, most studies focus on mature leukocyte populations and do not adequately address granulocytic maturation stages. Accurate identification of promyelocytes, myelocytes, metamyelocytes, band neutrophils, and segmented neutrophils is critical for assessing left shift, marrow response, leukemoid reactions, and hematological malignancies. This study aimed to develop and validate deep learning model for classification of peripheral blood cells with emphasis on granulocytic maturation stages.

**Methods:**
An open-source Kaggle dataset (MIT License) comprising 21,234 labeled peripheral blood cell images representing 11 hematopoietic cell classes was utilized. Images were divided into training (n=14,858), validation (n=3,185), and independent test (n=3,191) datasets. Image preprocessing included resizing to 224×224 pixels and augmentation using random horizontal flipping and rotation. A pretrained EfficientNet-B0 convolutional neural network was fine-tuned using the Adam optimizer and cross-entropy loss function. Model performance was assessed using accuracy, precision, recall, F1-score, and confusion matrix analysis.

**Results:**
The model achieved a training accuracy of 98.78%, with a peak validation accuracy of 95.10% at epoch 8 and an independent test accuracy of 94.33%. The weighted and macro F1-scores were 94.34% and 95.47%, respectively. Eosinophils and platelets achieved perfect classification performance (F1-score = 1.00), basophils, erythroblasts, monocytes, and lymphocytes demonstrated F1-scores exceeding 0.97. Most classification errors occurred between morphologically adjacent granulocytic maturation stages, particularly myelocytes and metamyelocytes, and between banded and segmented neutrophils. Misclassification between unrelated cell types was uncommon.

**Conclusion:**
EfficientNet-B0 demonstrated high-performance multiclass classification of peripheral blood cells, including clinically relevant granulocytic maturation stages. The biologically plausible error patterns observed support validity of the model and highlight its potential for AI-assisted hematological morphology assessment, automated differential counting, and decision-support applications. Keywords: Artificial Intelligence; Deep Learning; EfficientNet-B0; Granulocytic Maturation; Blood Cell Classification; Digital Hematopathology.

# 142 - Clinical Deployment and Ongoing Real-World Evaluation of an AI-Assisted Lymph Node Metastasis Detection Module in a National Reference Laboratory

**Rajiv Tangri[1], Gaurav Jain[1], Reena Nakra[1], Vandana Lal[1], Amala Abbas[2], Wai Po Kevin Teng[2], Yogesh Pratap Singh[2]**

[1] Dr Lalpath Labs

[2] Qritive Pte ltd

**BACKGROUND**

Accurate identification of nodal metastasis is critical as it directly influences cancer staging, prognosis and management decisions. However, the detection of metastasis remains challenging due to extensive manual search required and sometimes very small size. This study presents results of the real-world deployment of an Artificial intelligence (AI) assisted module for lymph node metastasis detection in clinical practice at Dr Lal Pathlabs' National Reference Laboratory, Delhi which caters to a very high volume of radical oncology specimens requiring screening of hundreds of lymph node section daily for detection of metastasis. DESIGN We used a segmentation based AI module to aid in efficient identification of lymph node metastasis. The module uses a CNN (Convolutional Neural Network) feature extractor that produces pixel-level classifications. It was trained on Whole slide images (WSIs) from lymph node biopsies (containing metastatic deposits from breast, colon, and gastric carcinoma). Following institutional deployment in September 2025, the module was integrated into routine diagnostic workflows. A retrospective and prospective real-world performance evaluation study using an expanded deployment dataset is being presented here.

**RESULTS**

In the initial phase of study, we were able to achieve a very high sensitivity of 99.0% and a very high negative predictive value of 99.4%. During clinical deployment various strategies were used and improvised with time following feedback from multiple users (digital histopathologist).

**CONCLUSION**

We have observed enhanced diagnostic efficiency and screening sensitivity in digital pathology workflows following use of this AI tool. The AI module showed rapid and consistent performance across multiple tumor types, supporting its utility as an assistive screening tool more so in a high volume laboratory set up. Ongoing real-world deployment study additionally highlights need for workflow optimization strategies to further improve reporting efficiency and operational utility in clinical practice.

# 143 - Weakly Supervised Multiple Instance Learning for BRAF Mutation Status Prediction in Thyroid Fine Needle Aspiration Cytology Using Cytology-Specific Foundation Models

**Hongjun Yoon[1], Yujin Lee[2], Seong Min Hong[3], Hyun-Joo Choi[2], DaeKy Jeong[1], Hyeongyeol Lim[1], Mohammad Rizwan Alam[3], Won June Cho[1], Yosep Chong[3]**

[1] Deepnoid Inc.
[2] St. Vincent's Hospital, College of Medicine, The Catholic University of Korea
[3] The Catholic University of Korea College of Medicine

**Background/Objective**

BRAF V600E mutation is a critical molecular alteration in thyroid carcinoma, informing diagnosis, prognosis, and targeted therapy selection. Conventional cytological evaluation, while offering excellent morphological assessment, cannot independently determine mutation status, necessitating separate molecular testing. We aimed to develop and validate an AI model leveraging weakly supervised learning to predict BRAF mutation status directly from thyroid fine needle aspiration (FNA) cytology whole slide images (WSIs), using only slide-level mutation labels without patch-level annotation.

**Methods**

A total of 263 thyroid FNA cytology WSIs (245 BRAF-positive, 18 BRAF-negative) were collected from two institutions and digitized with extended z-stacking. Features were extracted using four Vision Foundation Models (VFMs): CytoFM (cytology-specific), UNI2-h, UNI, and Virchow2. Clustering-constrained Attention Multiple Instance Learning (CLAM) was used for slide-level binary classification. To address severe class imbalance (positive-to-negative ratio 13.6:1), SMOTE-based feature interpolation, bag-level class weighting, and patch dropout were applied. Bayesian hyperparameter optimization was performed using stratified 5-fold cross-validation.

**Results**

CytoFM significantly outperformed all general-purpose histopathology VFMs, achieving a best AUROC of 0.784, compared to 0.715 for UNI2-h (non-overlapping 95% confidence intervals). Attention heatmaps demonstrated model focus on clinically relevant regions including dense cell clusters and atypical morphological features, while background and artifactual areas received low attention scores.

**Conclusion**

Weakly supervised deep learning can predict BRAF mutation status from thyroid FNA cytology WSIs with clinically meaningful accuracy. Cytology-specific pretrained foundation models substantially enhance performance over histopathology-trained counterparts. This approach holds promise as a non-invasive screening tool to triage molecular testing in indeterminate thyroid FNA cases, potentially reducing unnecessary testing burden and diagnostic delays.

# 144 - Deep learning for microsatellite instability in colorectal and gastric cancer: external validation across nine Korean hospitals

**Daeky Jeong[1], Rizwan Alam Mohammad[2], Won June Cho[1], Hongjun Yoon[1], Kwangil Yim[2], Kyung Jin Seo[2], Seongmin Hong[2], Hye Jin Ahn[2], Sung Hak Lee[2], Jisup Kim[3], Seung Yeon Ha[3], Ju-Yeon Pyo[4], Chan Kwon Jung[2], Hyun-Jong Jang[5], Yosep Chong[2]**

[1] Deepnoid
[2] Department of Hospital Pathology, College of Medicine, The Catholic University of Korea
[3] Department of Pathology, Gil Medical Center, Gachon University College of Medicine
[4] Department of Pathology, International St. Mary's Hospital, Catholic Kwandong University College of Medicine
[5] Department of Physiology, College of Medicine, The Catholic University of Korea

**Background/Objective**

Microsatellite instability (MSI) status guides immune-checkpoint inhibitor eligibility in colorectal (CRC) and gastric cancer (GC). Polymerase chain reaction (PCR) and immunohistochemistry remain the clinical standard but are resource-intensive and may be influenced by intratumoral heterogeneity due to single-block sampling. Deep learning (DL) from haematoxylin and eosin (H&E)-stained whole-slide images (WSIs) offers complementary screening potential. We assessed whether DL can both predict MSI and prioritize cases for repeat molecular testing in a multi-institutional dataset.

**Methods**

We assembled a retrospective cohort of 3,171 patients (CRC n=2,167; GC n=1,004) contributing 6,114 WSIs from nine Korean hospitals. External generalizability was assessed by an institution-based hold-out split (CRC: 4 training + 3 held-out hospitals; GC: 3 + 3). Two complementary classifiers were trained from a single MSI-High (MSI-H)/MSS label per patient: a patch-based convolutional neural network (CNN) operating on tumor patches and a clustering-constrained attention multiple instance learning (CLAM), with Vahadane stain normalization applied as preprocessing. AI-discordant MSI-H patients in the held-out test cohort underwent central PCR revalidation.

**Results**

In the CRC held out test cohort, the CNN-CLAM ensemble achieved patient-level area under the receiver operating characteristic curve (AUROC) 0·970 (CI 0·953-0·981); in the GC cohort, the ensemble achieved patient-level AUROC 0·835 (CI 0·729-0·923) with negative predictive value (NPV) 0·956. Vahadane stain normalization maintained AUROC ≥0·80 across all three external CRC hospitals, supporting robustness to inter-institutional staining variation. Central revalidation of four AI-discordant test-cohort patients reclassified three (75%) from MSI-H to MSS or MSI-low, in concordance with the model. Patient-level CLAM aggregation yielded low MSI-H probabilities (0·00-0·03) in all three reclassified cases, with attention concentrated on conventional well-differentiated tubular morphology consistent with the corrected labels.

**Conclusion**

DL from routine H&E both predicts MSI with high external accuracy and prioritizes candidate cases for repeat molecular testing, generalizing across the largest reported Korean MSI cohort.

# 145 - Clicks vs Corridors: Comparing Online and Onsite Digital Pathology Education Using Kirkpatrick's Model

**Nazish Jaffar[1], Mohsin Bilal[2]**
[1] Jinnah Sindh Medical University
[2] Prince Sattam Bin Abdulaziz University

**Objective**
We aimed to train cytologists/pathologists & evaluate their responses for two digital pathology educational interventions, an international online webinar and an onsite workshop at Karachi, Pakistan, using Kirkpatrick's model of medical education.

**Methods**
Two facilitators, an AI expert and a pathologist facilitated the training on QuPath software for tasks including creating folders, importing images, creating annotation labels, using different tools to annotate cells & tissues, performing cell detection, measurement analyses and exporting the results. Participants from a webinar titled 'Cytology AI Blueprint' (n=14 respondents) & an onsite hands-on workshop 'Harnessing health through AI' (n=13 respondents)were included. A comparative pre and post-intervention survey was conducted & responses were collected via online forms. We assessed Kirkpatrick Level 1 & 2 outcomes (satisfaction & engagement) and (knowledge & confidence) using Likert scale and dichotomous Yes/No items. Level 3 (intended application of skills) was also inquired. Wilcoxon Signed-Rank Test was applied for ordinal variables. Between-group comparison was performed using Mann-Whitney U Test.

**Result:**
About 70-80% participants from both sessions were not familiar with DP & QuPath. They reported high levels of satisfaction and engagement (Web: Median 3, IQR 1, Wshop: Median 4, IQR 1) on a 5 point Likert scale. Most respondents indicated improved understanding of digital pathology concepts (Web & Wshop: Median 3, IQR 0) and increased confidence in interpreting digital images and annotations. Participants from both webinar & workshop demonstrated a good perceived confidence in hands-on annotation skills (Web & Wshop: Median 3, IQR 1). Both groups appreciated the role of facilitators & suggested multiple sessions of longer duration. Between-group comparison revealed no statistically significant difference in post-intervention outcomes between webinar and workshop participants (p=0.259).

**Conclusion:**
Both webinar and workshop delivery models successfully enhanced digital pathology competency. This supports flexible, inclusive education strategies in pathology training worldwide and specially in LMICs.

# 147 - Artificial Intelligence–Based Tumor Stroma Ratio Assessment in Oral Squamous Cell Carcinoma: Validation Against Expert Consensus and Interobserver Variability Analysis

**Rajasa Jialdasani[1], Sujit Joshi[2], Abhijit Patil[3], Shanu Pratap[3], Gourav Agrawal[4], Sharvari Vishwasrao[5], Sangeeta Desai[6]**

[1] SenzanAI
[2] Excel Diagnostics
[3] Pixonate Labs
[4] Jupiter Hospital
[5] Independent Consultant
[6] Former Prof and Head, Tata Memorial Centre

**Background:**
Tumor-stroma ratio (TSR) is an emerging quantitative prognostic marker in oral squamous cell carcinomas (OSCC) with studies demonstrating that stroma rich tumors (High TSR) corelate with aggressive biology and worse survival. Manual TSR scoring can be affected by inter-observer variability. Advances in artificial intelligence (AI) in digital pathology can standardize TSR quantification.

**Objective:**
Evaluate the concordance of an AI-based TSR quantification system with expert pathologist consensus and to compare AI performance against interobserver variability as reference benchmarks.

**Methods:**
A retrospective cohort of OSCC cases with 98 regions of interest (ROIs) from invasive tumor front from whole-slide images were independently scored for TSR by three pathologists according to established TSR scoring guidelines. The same ROIs were analyzed using a deep semantic segmentation model trained to partition the tissue into clear structural compartments to generate continuous TSR measurements. Interobserver agreement among pathologists was assessed using correlation and agreement analyses. AI-derived TSR values were compared against the consensus pathologist assessment. Visual review of segmentation overlays was performed to assess biological plausibility of model outputs.

**Results:**
Interobserver agreement among pathologists was moderate -to-substantial (weighted kappa 0.69, Pearson r of 0.88 and Spearman p of 0.87. Against pathologist consensus, the model achieved Pearson correlation 0.88 and Spearman correlation 0.88. Overall, 72.4%, 82.7% and 93.9% of ROIs were within ±10, ±15 and ±20 TSR scores of consensus. Bland-Altman analysis showed a modest positive bias of +4.58 TSR points, indicating mild model overestimation. Review by pathologists of the visual overlays revealed satisfactory segmentation with minor discrepancies noted in 5 ROIs.

**Conclusion:**
AI-based TSR quantification demonstrates high concordance with expert assessment. By reducing observer variability and enabling standardized quantitative evaluation of the tumor microenvironment, AI-assisted TSR assessment may facilitate broader adoption of this prognostic biomarker within routine digital pathology workflows and future risk stratification models for OSCC.

# 149 - Comparative Evaluation of Classical and Quantum Machine Learning for Binary Discriminator between Neutrophils and Lymphocytes

**Parikshit Sanyal, Ankur Ahuja, Venkatesan Somasundaram**
AFMC Pune

Introduction & Rationale This study outlines a controlled, reproducible comparative framework between classical machine model and emerging Variational Quantum Classifiers (VQCs) using Hybrid Quantum-Classical Neural Networks (HQCNNs) to differentiate between neutrophils and lympohcytes. The objective is to assess whether quantum-enhanced feature mapping provides any advantage in accuracy, time for training and performance, as well as representation of classes in quantum bits.

**Methods**
An open source dataset of blood cell images (from Kaggle, under CC by SA license 4.0, https://www.kaggle.com/datasets/unclesamulus/blood-cells-image-dataset) was used to train a classical Deep Convolutional Neural Network. In addition, a HQCNN was also trained on the same dataset. The performance of the two models were compared. The quantum ML model was configured for angle embedding, i.e. reduced classical feature vector was mapped onto an qubit register. Each feature parameterized a single-qubit rotation gate around the Y-axis. This component is responsible for encoding the classical input data into quantum states. A Strongly Entangling Circuit layout was repeated for quantum layers. Each layer contains parameterized rotation gates for every qubit, followed by a linear chain of Controlled-NOT (CNOT) gates to induce multi-qubit quantum entanglement across the cellular feature space. The qubits were finally measured.

**Results**
The time taken for training over 50 epochs was 6.74 times (3183 seconds vs 472 second) higher for the QML model on a standard AER simulator (using the QisKit implementation). The final test accuracy in the hybrid model was 71.43%, vs 99% in classical model.

**Conclusion**
Although the performance of hybrid model was not nearly comparable to the classical model, analysis of the Bloch sphere representing cubits of the two classes displayed how the two classes are represented in quantum states. This might provide novel encoding mechanisms for higher dimensional data, such as those met in pathology practice, in quantum computing era.

# 151 - RELIABILITY OF QUPATH-BASED DIGITAL IMAGE ANALYSIS IN ASSESSING CD138, MUM1, AND KI67 IHC EXPRESSION IN MULTIPLE MYELOMA

**Swesti sari Suciati, Indrawati Indrawati, Ery Kus Dwianingsih**
Department Of Anatomical Pathology, Faculty Of Medicine, Public Health, And Nursing, Universitas Gadjah Mada

**Background:**
Immunohistochemical (IHC) evaluation of CD138, MUM1, and Ki67 in multiple myeloma (MM) has been performed using semiquantitative manual scoring with inherent interobserver variability and reproducibility issues, especially with the heterogeneous plasma cell distribution in bone marrow. Digital pathology based on Whole Slide Image (WSI) with QuPath software offers an objective and quantitative alternative for IHC evaluation, but its application and validation in MM specimens in Indonesia is underexplored. The aim of this study was to describe the expression profiles of CD138, MUM1, and Ki67 in MM patients using WSI-based digital pathology and QuPath quantification, and to evaluate the reliability of this approach compared to conventional manual assessment.

**Method:**
CD138 expression was high and relatively homogeneous, with comparable means between manual and QuPath assessment (90.56 ± 12.41% vs 90.07 ± 14.61%, n=72), showing good agreement (single-measures ICC=0.630, 95% CI: 0.467–0.752; average-measures ICC=0.773, 95% CI: 0.637–0.858; $p<0.001$). MUM1 expression showed greater variability between methods (81.62 ± 26.80% vs 78.31 ± 28.57%, n=52), with good agreement (single-measures ICC=0.729, 95% CI: 0.573–0.835; average-measures ICC=0.843, 95% CI: 0.728–0.910; $p<0.001$). Ki67 demonstrated the highest concordance (31.50 ± 29.56% vs 31.71 ± 29.66%, n=14), with excellent agreement (single-measures ICC=0.955, 95% CI: 0.864–0.985; average-measures ICC=0.977, 95% CI: 0.927–0.993; $p<0.001$).

**Conclusion:**
WSI-based digital pathology using QuPath offers a reliable and reproducible method for quantifying CD138, MUM1, and Ki67 expression in multiple myeloma, demonstrating good-to-excellent agreement with manual scoring. The highest concordance was observed for Ki67, supporting digital quantification as a standardised approach for IHC assessment in MM diagnostic and research workflows. Keywords: Multiple myeloma, CD138, MUM1, Ki67, digital pathology whole slide image, QuPath.

# 153 - Use of Artificial Intelligence based deep learning model in identification of nucleated red blood cells on digital images in low-resource settings

**Maryam Raza**[1]
[1] National Institute Of Child Health
[2] Jinnah Sindh Medical University
[3] Artificial Intelligence Technology centre

**Background**
The detection of nucleated red blood cells (NRBCs) in peripheral blood smears is crucial in diagnosis of certain systemic and hematological disorders. It is also important to calculate the number of NRBC, to correct the falsely high leukocyte count. Conventional method of examining peripheral smears is time consuming and tedious. Deep learning-based algorithms have secured significant attention due to their ability to precisely deal with intricate visual data and assist pathologists vigilantly. This research authenticates the use of artificial intelligence based deep learning models in the identification of NRBCs from digital images of peripheral smears. Material and methods Prospective cross-sectional study was conducted in hematology laboratory of National Institute of Child Health. Digital images of selected NRBC pathology cases were captured through a camera connected microscope at 40x. Out of 803 images,695 were positive and 108 were negative. Images were annotated for variety of NRBC by a hematologist. AI based Deep learning models MedSAM2, YOLOv11n-seg, and Mask R-CNN were evaluated for NRBC detection and segmentation by computational pathologist. Model performance was assessed using sensitivity, specificity, and F1-score on training, validation, and test datasets.

**Results**
The data was split into testing, training and validation, where MedSAM2 achieves the highest overall performance with test F1 ≈ 0.876, followed by YOLOv11n-seg with F1 ≈ 0.801, and Mask R-CNN with F1 ≈ 0.74. The sensitivity and specificity of Med SAM2 was 88.56% and 87.34% respectively.

**Conclusion**
Overall, this small study predicts the potential to deep learning models applicability at low resource setting. As MedSAM2, which have strong potential to assist NRBC detection in hematology practice , which can further be implicated in diagnosis of other types of crucially significant cells in diagnostics like malarial parasites, blast or atypical mononuclear cells in peripheral smear for vigilant response and action.

# 154 - Digital Pathology Adoption and Readiness in Malaysia: A Cross-Sectional Survey of a Middle-Income Country

**Zahrah Tawil[1], Elaine Wan Ling Chan[2,4], Shew Fung Wong[3,4], Zanariah Alias[5], Salmi Abdullah[1], Inti Zlobec[6,7], Lai Meng Looi[8]**

[1] Hospital Selayang, Ministry Of Health Malaysia
[2] Department of Data Science and Artificial Intelligence, Faculty of Engineering and Technology, Sunway University
[3] School of Medicine, Department of Pathology and Pharmacology, IMU University
[4] Institute for Research, Development and Innovation, IMU University
[5] Department of Pathology, Hospital Kuala Lumpur
[6] Institute of Tissue Medicine and Pathology, University of Bern, Bern, Murtenstrasse 31
[7] Department of Digital Medicine, University of Bern, Murtenstrasse 31
[8] Department of Pathology, Faculty of Medicine, University of Malaya

**Introduction**

Digital pathology (DP) enables whole-slide imaging, remote reporting, data sharing, and integration of artificial intelligence (AI), with benefits for diagnostic accuracy, workflow efficiency, education, and research. National data on DP adoption, readiness, and stakeholder perceptions in Malaysia remain limited. This study evaluated DP adoption and readiness among Malaysian pathologists and related stakeholders. Material and methods A cross-sectional survey of 56 professionals, including pathologists and non-pathologist healthcare stakeholders, was conducted to assess demographics, professional experience, DP familiarity, implementation stage, readiness to use DP and AI-assisted tools, perceived benefits and barriers, and confidence in national readiness by 2030. Quantitative data were analysed descriptively and with Pearson correlation; qualitative responses were thematically analysed.

**Results**

Regional engagement varied, with Central Malaysia (Kuala Lumpur/Selangor) contributing most respondents. Awareness of DP was high, but implementation was mainly in early adoption or developmental phases. Pathologists strongly agreed that DP enhances diagnostic accuracy and residency training, with moderate concern about skill erosion and minimal concern about AI-related detraining. Perceived readiness to use DP was strongly correlated with readiness to adopt AI-assisted tools ($r=0.75$), whereas years of professional practice and self-reported familiarity with DP were each only weakly associated with readiness to use DP and AI-assisted tools. Keybarriers included costs, infrastructure limitations, workforce gaps, and regulatory uncertainty; benefits centred on efficiency, remote reporting, and support for education and research.

**Conclusion**

DP adoption in Malaysia is at an early stage. Targeted investments in infrastructure, training, funding, public–private collaboration, and policy development are essential to accelerate adoption. Keywords: Digital Pathology, Malaysia, Implementation Readiness, AI Adoption

# 155 - IDH1 Mutation Prediction in Cholangiocarcinoma from Whole-Slide Images Using Attention-Based Multiple Instance Learning with Confidence-Targeted Synthetic Augmentation

**Swapnil Bhat[1], Nikita Mulchandani[2], Roshani Gala[2], Devansh Lalwani[1], Jay Mehta[2], Maulik Shah[1]**
[1] Turocrates.ai
[2] Neuberg Oncopath

**Background:**
IDH1 mutations occur in roughly 10 to 13 percent of intrahepatic cholangiocarcinomas (IDH1/2 alterations up to about 20 percent), defining a subtype sensitive to FDA-approved IDH inhibitors. Confirming IDH status requires sequencing, which is slow, costly, and often unavailable. Predicting it from H&E whole-slide images could enable rapid, low-cost stratification at diagnosis. The obstacle is the scarcity of cholangiocarcinoma cohorts with paired histology and molecular labels, compounded by low mutant prevalence and cross-institution domain shift.

**Methods:**
We assembled 63 diagnostic WSIs from two institutions: 39 from TCGA-CHOL and 24 external, totaling 18 IDH-mutant and 45 wildtype. Tissue was segmented with GrandQC, Macenko-normalized, and tiled into 256-pixel patches at 20x. Patch embeddings used Virchow2, with UNI2-h and mSTAR for comparison. A gated-attention MIL classifier (CLAM-SB) was trained under 5-fold patient-disjoint cross-validation, alongside ABMIL and Feather. To address minority-class scarcity we synthesized tiles with PixCell, conditioned on the highest-attention mutant tiles. We also evaluated external oversampling and decorrelated-model ensembling.

**Results:**
The single-model baseline reached held-out AUROC 0.71 pooled and 0.81 on TCGA; accuracy (about 81 percent) tracked class prevalence, making AUROC and NPV the meaningful endpoints. Diffusion-based augmentation did not improve discrimination and slightly degraded it (AUROC 0.69), but reduced train-validation loss divergence, acting as a regularizer. The decisive gains came from ensembling decorrelated models, reaching pooled AUROC 0.79 and TCGA 0.91, with NPV 0.96 at 0.94 sensitivity. External performance remained the bottleneck (AUROC about 0.59), improving to 0.64 with oversampling. A transfer experiment revealed an architecture-robust asymmetry: the external-trained model generalized to TCGA (AUROC 0.68), whereas the TCGA-trained model failed externally (AUROC 0.54, sensitivity 0.00).

**Conclusion:**
Weakly-supervised MIL on foundation-model embeddings predicts IDH status with clinically meaningful rule-out performance (NPV 0.96) despite a small cohort. Synthetic augmentation regularized but added no discriminative signal; gains arose from ensembling and rebalancing.

# 156 - Bridging the Gap: Pathologist and Clinician Perspectives on Synoptic Reporting in Oncopathology in India

**Varun Savant[1], Swapnil Bhat[2], Devansh Lalwani[1,2], Maulik Shah[2], Ameya Rojekar[1], Pradeep Vaideeswar[1]**
[1] Seth Gsmc & Kem Hospital
[2] Turocrates.ai

**Background:**
Synoptic reporting improves completeness and reduces ambiguity in oncopathology reports compared with narrative formats and is embedded in international guidelines from the College of American Pathologists (CAP), the International Collaboration on Cancer Reporting, and the Royal College of Pathologists. Synoptic reports inform staging, adjuvant therapy planning, and tumour board discussions. However, adoption across Indian practice remains uneven, and the perspectives of both pathologists and treating clinicians on its utility, barriers, and facilitators have not been systematically characterized.

**Methods:**
A cross-sectional, questionnaire-based observational study was conducted. A structured instrument spanning seven domains was distributed via Google Forms to pathologists, surgical oncologists, medical oncologists, and radiation oncologists across teaching, tertiary, corporate, and standalone laboratory settings in India. The questionnaire comprised participant demographics, shared Likert-scale items, role-specific modules, a digital and AI dimension, barriers and local context items, and open-ended questions. Branching logic ensured role-appropriate routing. Convenience sampling through professional networks recruited 50 pathologists and 50 clinicians.

**Results:**
Among pathologists, 47% cited insufficient training and poor subspecialty template fit as primary barriers, while only 27% identified time constraints. Nearly all pathologists (93%) preferred CAP guidelines. Key facilitators included electronic templates (73%) and mandatory institutional policy (67%); only 40% strongly agreed on current benefits. Among clinicians, 80% preferred synoptic over narrative formats, 73% reported difficulty extracting information from narrative reports, and 87% believed synoptic reporting would improve clinical decision-making. Essential elements identified included tumour type, lymph node status, margins, and TNM staging.

**Conclusion:**
Both pathologists and clinicians recognize the value of synoptic reporting, but implementation is hindered by training gaps and template limitations. Targeted training programs, improved electronic template design, and institutional mandates may accelerate adoption across Indian oncology practice.

# 158 - Automated Banff Lesion Scoring in Kidney Transplant Biopsies Using Foundation Model Features and Attention-Based Multiple Instance Learning

**Devansh Lalwani[1,2], Ameya Rojekar[1], Varun Savant[1], Swapnil Bhat[2], Maulik Shah[2]**
[1] Seth Gsmc & Kem Hospital
[2] Turocrates.ai

**Background:**
The Banff classification provides standardized histological scoring for kidney transplant rejection, assessing interstitial inflammation (i), tubulitis (t), glomerulitis (g), and intimal arteritis (v). Banff scoring is labor-intensive, subject to inter-observer variability, and requires nephropathology expertise that is scarce globally. Automated scoring from digitized biopsies could improve reproducibility and access. Prior computational approaches have relied on compartment-specific segmentation, whereas slide-level weakly-supervised methods remain underexplored for ordinal Banff scores.

**Methods:**
We utilized 81 PAS-stained kidney transplant WSIs from the MONKEY Challenge dataset with expert Banff scores. Patch-level features (768-dimensional) were extracted at 10x and 20x using the frozen Phikon foundation model. An ABMIL classifier with ordinal CORAL loss was trained under 5-fold cross-validation, evaluated by quadratic weighted kappa (QWK). Two architectures were compared: v1 baseline (1.32M parameters, embedding dimension 512) and v2 regularized (166K parameters, embedding 128, dropout 0.5, learning rate 3e-4). An ensemble with 20-fold test-time augmentation produced final predictions. For expert validation, 18 slides were selected via MC Dropout uncertainty sampling for blinded pathologist rescoring.

**Results:**
The ensemble achieved mean QWK of 0.243 on learnable scores (i, g, v), with individual scores of i=0.124, g=0.398, and v=0.208. The t-score was unlearnable (QWK=0.000) due to extreme imbalance (only 2 positive cases among 81 WSIs). The regularized v2 model outperformed v1 on g-score (+0.052) and v-score (+0.084) despite 8-fold fewer parameters, demonstrating that compact heads better suit foundation model features. Conventional approaches (logistic regression, gradient boosting) on pooled features substantially underperformed ABMIL (mean QWK approximately 0.08).

**Conclusion:**
Weakly-supervised ABMIL with foundation model features provides a viable baseline for automated Banff scoring, with glomerulitis achieving the strongest agreement. Performance is constrained by limited training data rather than architecture, as the smaller regularized model proved superior. Expanding training cohorts and incorporating compartment-aware attention represent the most promising directions for clinically actionable agreement.

# 159 - End-to-End Automated Ki-67 Quantification Pipeline for Invasive Breast Cancer on Whole-Slide Images

**Himanshi Diwan[2], Gurudutt Gupta[2], Swapnil Bhat[3], Devansh Lalwani[1], Maulik Shah[3]**

[1] Turocrates.ai

[2] RGCIRC

[3] Turocrates.ai

**Background:**

The Ki-67 labelling index is a key prognostic biomarker in invasive breast cancer, but manual estimation suffers from poor inter-observer reproducibility and inconsistent region selection. Inclusion of in-situ or benign epithelium and Ki-67-positive lymphocytes inflates the index and undermines clinical utility. We present an end-to-end pipeline chaining tissue segmentation, invasive-tumour detection, cell classification and Ki-67 computation into a single automated whole-slide-image (WSI) workflow.

**Methods:**

The pipeline comprises four stages. (1) A deep-learning model separates tissue from background. (2) An invasive-tumour detector, using Virchow2 foundation-model embeddings with a lightweight segmentation head trained on pathologist-annotated regions, localises invasive carcinoma while excluding in-situ disease, benign epithelium and stroma. (3) Within invasive regions, nuclei are detected by an inverted-contrast method and described by classical morphometric features plus a slide-relative size descriptor encoding that tumour nuclei are consistently larger than neighbouring lymphocytes and stroma; a random-forest classifier labels each nucleus tumour versus non-tumour. Ki-67 positivity is read per nucleus from DAB intensity via Macenko deconvolution. (4) The index is computed as positive-over-total tumour nuclei across the invasive compartment, with hotspot tiles ranked by proliferation, consistent with St Gallen-style scoring.

**Results:**

The invasive-tumour detector achieved a slide-stratified held-out AUROC of 0.78. Restricting scoring to classifier-confirmed tumour nuclei raised the Ki-67 index from 48% (ungated) to 60% (tumour-gated); the excluded non-tumour population retained a 40% labelling rate, confirming the gate removes a genuine lower-proliferation population rather than imposing a size-threshold artefact. Each detected cell is returned with spatial coordinates and classification label, enabling direct overlay onto the WSI for pathologist verification.

**Conclusion:**

This pipeline automates Ki-67 quantification exclusively within invasive tumour, addressing a principal source of scoring variability and providing an interpretable, scalable solution for routine surgical pathology. Validation against pathologist consensus scoring is ongoing.

# 160 - Foundation Model-Based Prediction of Microsatellite Instability in Endometrioid Endometrial Cancer

**Aekta Shah[2], Santosh Menon[2], Devansh Lalwani[1,3], Swapnil Bhat[3], Maulik Shah[3], Rishi Balse[1]**

[1] Seth Gsmc & Kem Hospital
[2] Tata Memorial Hospital
[3] Turocrates.ai

**Introduction:**
Microsatellite instability (MSI) from mismatch repair deficiency is a clinically relevant biomarker in endometrial cancer, informing immune checkpoint inhibitor eligibility. Routine MSI testing is constrained by cost and laboratory resources. Deep learning predicts MSI from H&E histology in colorectal cancer, but foundation model application to endometrioid endometrial cancer has not been systematically studied.

**Methods:**
H&E whole slide images from the TCGA-UCEC cohort were restricted to endometrioid histology with confirmed MSI status (MSIsensor ≥ 10 for MSI-H; < 4 for MSS), yielding 348 patients (115 MSI-H, 233 MSS). Six foundation models (UNI, CONCH, Virchow, Phikon-v2, GigaPath, PLIP) served as fixed feature extractors. Two strategies operating on different inputs were compared: slide-level weakly supervised MIL (ABMIL, TransMIL, CLAM-SB) using the entire WSI, and patch-level classical classifiers (SVM, Logistic Regression, Random Forest, KNN) with majority voting, operating on tumor patches selected by the pathologist rather than the full slide. Performance was assessed by 5-fold patient-stratified cross-validation using AUC, accuracy, weighted precision, recall, and F1.

**Results:**
Among MIL architectures, CLAM-SB performed best (AUC 0.77; accuracy 0.74), followed by ABMIL (AUC 0.70) and TransMIL (AUC 0.65), using only slide-level labels. For patch-level classification, Virchow embeddings outperformed Phikon-v2 across classifiers. Virchow with Random Forest achieved the best overall result (accuracy 0.83, weighted precision 0.85, F1 0.81), with Virchow plus Logistic Regression comparable (accuracy 0.82). KNN was weakest throughout. The two frameworks differ in input granularity yet reach comparable MSI prediction.

**Conclusion:**
Foundation model feature extraction enables MSI prediction in endometrioid endometrial cancer under both slide-level and patch-level frameworks, with Virchow embeddings superior. The patch-level approach derives predictions from pathologist-selected tumor patches rather than reconstructed whole slides.

# 161 - Development of an AI-Based Bone Marrow Differential Cell Count System: A Retrospective Validation Study

**Devansh Lalwani[1], Nikita Mehra[2], Anuradha Talukdar[2], Swapnil Bhat[3], Maulik Shah[3]**
[1] Seth Gsmc & Kem Hospital
[2] CCHRC
[3] Turocrates.ai

**Background:**
Manual bone marrow differential cell counting is labor-intensive, time-consuming, and subject to significant interobserver variability, particularly for morphologically similar cell lineages. Automating this process could improve reproducibility and throughput in hematopathology laboratories. This study describes the development and retrospective validation of an end-to-end AI pipeline for automated bone marrow aspirate differential cell counting using whole-slide images.

**Methods:**
The AI system comprises three modules: (1) automated region-of-interest selection to identify bone marrow particles and optimal monolayer counting areas, (2) cell segmentation using Cell-SAM, a segment-anything model adapted for cellular morphology, and (3) cell classification using multi-classifier feature extraction. Retrospective validation was performed on over 200 archived Wright-Giemsa and Leishman-stained bone marrow aspirate whole-slide images. Expert hematopathologists performed blinded 500-cell manual differential counts as the reference standard. The AI system classified cells into 14 hematopoietic categories including myeloid precursors, erythroid precursors, lymphocytes, plasma cells, monocytes, eosinophils, basophils, and blasts.

**Results:**
The AI system achieved 85% overall accuracy with balanced accuracy of 85% across all 14 categories. Top-performing categories included eosinophils (precision 98.6%, recall 94.7%), plasma cells (precision 91.6%, recall 87.3%), and erythroblasts (precision 87.0%, recall 89.3%). The automated pipeline analyzed substantially more cells than the 500-cell manual standard in a fraction of the time, improving statistical robustness of differential counts while reducing turnaround time and labor requirements.

**Conclusion:**
This AI-based bone marrow differential counting system demonstrated strong performance across 14 hematopoietic cell categories, with particularly high accuracy for eosinophils, plasma cells, and erythroblasts. The system may reduce workload and interobserver variability in hematopathology. Prospective multi-center validation is warranted to establish clinical utility.

# 163 - Comparing Scanner-Based and Microscope-Based Workflows for Ki-67 Quantification

**Devansh Lalwani[1], Jay Mehta, Nikita Mulchandai[2], Roshani Gala[2], Sheena Alphones[3], Swapnil Bhat[1], Varun Savant[4]**

[1] Turocrates.ai
[2] Neuberg Oncopath
[3] NM Medical
[4] Seth GSMC & KEM Hospital

**Background:**
The Ki-67 labelling index assesses tumor proliferation in breast cancer and informs treatment decisions including adjuvant chemotherapy selection. Manual Ki-67 counting is time-consuming and subject to interobserver variability. Scanner-based digital pathology workflows with dedicated image analysis software are increasingly adopted, but high infrastructure costs constrain accessibility in many low- and middle-income country (LMIC) laboratories. We developed and fine-tuned a segmentation model for tumor patches to exclude spindle cells and lymphocytes, enabling Ki-67 quantification from images in which multiple regions are photographed or a video is recorded over the same tumor area.

**Methods:**
Fifty breast cancer patients with over 500 patches at 20x magnification were analyzed. Two experienced pathologists independently assigned Ki-67 percentages as the reference standard. The scanner-based workflow used a whole-slide imaging scanner with bundled proprietary analysis software. The microscope-based workflow captured identical regions of interest using a camera attached to a trinocular microscope port. The two workflows differ in acquisition hardware while operating on the same regions of interest. Performance was assessed using Pearson correlation, mean absolute error (MAE), and Bland-Altman analysis.

**Results:**
Inter-pathologist agreement was strong (r=0.96, MAE=5.0%). The scanner-based workflow showed excellent correlation (r=0.938-0.965) but consistently overestimated Ki-67 by 7-10 percentage points (MAE=7.6-10.5%). The microscope-based workflow achieved comparable correlation (r=0.919-0.954) with lower bias (+1.2-4.2 percentage points) and lower MAE (5.3-6.0%). The microscope-based approach reduced infrastructure costs by 98% and operated without dedicated technical personnel.

**Conclusion:**
A microscope-based Ki-67 quantification workflow reached accuracy comparable to or better than the scanner-based approach, with lower systematic bias and 98% lower infrastructure cost. The microscope-based workflow derives Ki-67 indices from camera-captured regions of interest rather than whole-slide scans, a mode of acquisition that does not depend on whole-slide scanner infrastructure.

# 165 - AI-Assisted Endometrial Cycle Phase Dating: Model Development and Retrospective Workflow Validation at a Tertiary Care Centre

**Anshu Shetty[2], Gwendolyn Fernandes[2], Devansh Lalwani[1,2], Swapnil Bhat[1]**
[1] Turocrates.ai
[2] Seth GSMC & KEM Hospital

**Background:**
Endometrial cycle phase dating remains subjective and poorly reproducible between observers. Most computational approaches use slide-level multiple-instance learning with weak labels and report development metrics only, with little evidence on whether such a model fits real reporting practice. We developed a polygon-supervised patch classifier and then tested on consecutive cases at a tertiary centre.

**Methods:**
Fifty endometrial whole-slide images reported at a single tertiary care centre formed the retrospective validation set. The model was developed on the iCAIRD endometrial dating cohort and annotated endometrial WSIs (0.25 um/px, 40x), with pathologist-drawn polygon annotations in GeoJSON. An 8-class taxonomy (Proliferative, Secretory, Menstrual, Hormonal, Atrophic, Adenocarcinoma, Hyperplasia with Atypia, Insufficient) was coarsened to 5-class and 4-class (Benign, Malignant, Atypical, Insufficient) levels. Patches (256 x 256 at 20x) were labelled by polygon intersection (minimum 50% overlap). Frozen UNI2 features fed a dual-head architecture trained jointly: a patch head for dense per-tile supervision and an attention-based MIL head for slide-level prediction, with inverse-frequency sampling and focal loss for class imbalance. Development used patient-level 5-fold cross-validation with a held-out test split. The model was then locked and applied, unmodified, to the 50 tertiary-centre WSIs, scanned locally and never seen in development. The validation endpoint was agreement between model phase call and original sign-out diagnosis, with discordances re-reviewed by a masked pathologist. Tile-resolution class maps were assessed for interpretability.

**Results:**
On held-out development data the locked model reached a slide-level AUROC of 0.92 with comparable accuracy across phase classes. Retrospective validation on the 50-slide set is in progress; concordance with sign-out, per-class agreement, discordance outcomes, and class-map assessment will be reported.

**Conclusion:**
Results will indicate whether this approach suits routine endometrial dating and grading of endometrial adenocarcinoma

# 167 - Weakly-Supervised Detection of Helicobacter pylori Infection from H&E-Stained Gastric Biopsies Using Foundation Model Features

**Sheena Alphones[2], Zeeshanali Fazalbhoy[2], Devansh Lalwani[1], Swapnil Bhat[1], Maulik Shah[1]**
[1] Turocrates.ai
[2] NM Medical

**Background:**
Helicobacter pylori infection is a major risk factor for gastric carcinoma and peptic ulcer disease, affecting over half the global population. Histopathological detection typically relies on special stains (Giemsa, Warthin-Starry) or immunohistochemistry, adding cost and turnaround time. Detection from routine H&E biopsies alone is challenging due to the organism's small size, yet would eliminate ancillary staining in screening workflows. Prior deep learning approaches have included special-stain slides, risking stain-type classification shortcuts rather than learning genuine morphological features.

**Methods:**
We curated 140 H&E-stained gastric biopsy WSIs (111 H. pylori-positive, 29 negative) from 77 patients in the HISTAI gastrointestinal dataset. Non-H&E slides (39 Giemsa, 3 Congo Red, 2 Desmin/SMA) were excluded to prevent stain-type shortcuts. Patch-level features (2560-dimensional) were extracted using the Virchow2 foundation model. A CLAM-SB architecture with gated attention and instance-level clustering regularization was trained for slide-level classification using stratified 5-fold cross-validation. The model employed a 512-dimensional intermediate representation with BCE and instance discrimination loss. Real-life clinical validation is ongoing, with prospective cases evaluated against paired special-stain ground truth.

**Results:**
The model achieved mean AUC of 0.717 +/- 0.099 across five folds (per-fold: 0.764, 0.875, 0.604, 0.722, 0.622). The best fold achieved AUC of 0.875, accuracy of 90.0%, sensitivity of 92.3%, specificity of 75.0%, and F1-score of 0.941. High fold-wise variance reflects severe class imbalance (79% positive) and few negative cases (approximately 6 per validation fold). Attention maps highlighted surface epithelial and foveolar regions, aligning with known H. pylori colonization sites.

**Conclusion:**
Ongoing clinical validation will establish generalizability as a screening tool to triage cases for confirmatory staining, reducing costs and turnaround time in gastroenterology services.

# 168 - Predicting Tubercular Etiology of Granulomatous Inflammation from H&E-Stained Sections: A GeneXpert-Validated Triage Study

**Sheena Alphones[1], Zeeshanli Fazalbhoy[1], Devansh Lalwani[2], Swapnil Bhat[2], Maulik Shah[2]**
[1] NM Medical
[2] Turocrates.ai

**Background:**
Granulomatous inflammation on hematoxylin-and-eosin (H&E) sections encompasses a broad differential including tuberculosis, sarcoidosis, fungal infections, and foreign body reactions. In TB-endemic regions such as India, confirming mycobacterial etiology requires GeneXpert (Xpert MTB/RIF) molecular testing, adding cost and turnaround time. Morphologic features on H&E, including caseous necrosis pattern, giant cell morphology, and inflammatory composition, may carry latent predictive signal for TB positivity. We present an AI-based triage tool that predicts tubercular etiology directly from H&E sections to prioritise cases for confirmatory GeneXpert testing.

**Methods:**
The study uses two cohorts: a retrospective open-source dataset with paired GeneXpert ground truth for model development, and a retrospective reference-laboratory cohort for external validation. Case identification uses a locally deployed large language model to extract diagnostic terms from pathology reports, categorise cases, and pull the corresponding whole-slide images. The pipeline comprises tissue segmentation, patch extraction at 20x, feature embedding using a histopathology foundation model, and aggregation via a multiple-instance learning (MIL) classifier. Performance is reported as AUROC, sensitivity, specificity, and F1. Predictions are benchmarked against board-certified pathologist visual assessments of TB likelihood.

**Results:**
On the open-source development cohort, the MIL classifier achieved an AUROC of 0.85 for distinguishing TB-positive from TB-negative granulomas on H&E alone. External validation on the reference-laboratory cohort is currently underway. Feature analysis highlights caseous necrosis extent and perilesional lymphocytic composition as discriminative morphologic patterns.

**Conclusion:**
An H&E-based triage tool for tubercular granulomatous inflammation could accelerate diagnosis in resource-constrained, TB-endemic settings.

# 172 - Digital Pathology and Deep Learning in Sebaceous Tumors: Morphometric Comparison of Adenoma and Carcinoma

**Rabia Qureshi[1], Nazish Jaffar[2]**
[1] Aga Khan University Hospital
[2] Jinnah Sindh Medical University

**Background:**
Cutaneous sebaceous neoplasms are extremely rare in histopathology practice. It is essential to diagnose them accurately because of a considerable morphological overlap with inflammatory skin lesions and other skin tumors, and an association with Muir-Torre syndrome. The study was conducted to evaluate the utility of open-source digital pathology software for quantitative morphometric analysis and comparison of nuclear features in sebaceous adenoma and sebaceous carcinoma.

**Methods:**
Ten sebaceous adenomas and sebaceous carcinomas were each retrospectively selected. The H&E-stained photomicrographs were analyzed using QuPath version 0.7.0. Representative tumour regions were annotated manually, and automated cell detection using standardized parameters across all images was performed. Nuclear features, including mean nuclear area and mean nuclear circularity, were extracted and analysed using Microsoft Excel. Quantitative measurements were also compared between the two types of neoplasms.

**Results:**
The average number of cells detected (Num Detections) were 1487.6 ± 824.17 and 2200.5 ± 853.57 in sebaceous adenoma and carcinoma categories respectively. Quantitative morphometric analysis demonstrated a mean nuclear area of 160.39 ± 19.47 µm² in sebaceous adenomas compared with 150.78 ± 44.00 µm² in sebaceous carcinomas ($p = 0.90$). Mean nuclear circularity was 0.67 ± 0.007 in adenomas and 0.70 ± 0.027 in carcinomas ($p = 0.07$). Average N:C ratio was 0.38 ± 0.01 for benign lesions while it was 0.38 ± 0.02 for their malignant counterparts ($p = 0.96$). Sebaceous carcinomas exhibited larger nuclei and greater nuclear irregularity compared with sebaceous adenomas however there were no significant differences between their N:C ratios.

**Conclusions:**
This pilot study determines the feasibility of using QuPath for quantitative morphometric analysis of sebaceous neoplasms. Larger studies are required to validate the observed findings and explore their diagnostic utility, which will ultimately lead to correct treatment and better prognostic outcomes for the patients.

# 173 - Low-Cost Digital Pathology for Low- and Middle-Income Countries: A Scoping Review of Open-Source Imaging, AI, and Telepathology

**Pasindu Nanayakkara[1], Ozben Yalcin[2], Shanthi Velusamy[3], Ruseik Rahumath[4], Thiyaphat Laohawetwanit[5], Madhura Kulkarni[6], Yosep Chong[7]**

[1] Postgraduate Institute Of Medicine, University Of Colombo, Sri Lanka
[2] Department of Pathology, Cemil Tascioglu City Health Application and Research Center, University of Health Sciences
[3] Department of Pathology, Sri Shankara Cancer Hospital and Research Centre
[4] Health Information Management Unit, The National Hospital of Sri Lanka
[5] Division of Pathology, Chulabhorn International College of Medicine, Thammasat University
[6] Center for Translational Cancer Research, Prashanti Cancer Care Mission
[7] Department of Hospital Pathology, Uijeongbu St. Mary's Hospital, College of Medicine, The Catholic University of Korea

**Background:**
Digital pathology has transformed diagnostic workflows in high-income countries, yet adoption in low- and middle-income countries (LMICs) remains limited because of high equipment costs and workforce shortages. Recent advances in open-source hardware, artificial intelligence (AI), telepathology, and cloud-based platforms have created new opportunities for affordable digital pathology implementation. This scoping review maps the current landscape of low-cost digital pathology solutions applicable to LMIC settings.

**Methods:**
A scoping review was conducted in accordance with the PRISMA-ScR framework. Databases were searched from April 2018 to April 2026. PubMed/MEDLINE, Web of Science, IEEE Xplore, Google Scholar, conference proceedings, and open-source software repositories were searched for digital pathology solutions relevant to resource-limited settings. Eligible technologies were open-source, freely available, cost less than USD 5,000, demonstrated deployment or validation in LMICs, or were specifically designed for resource-constrained environments. Data were extracted on functionality, cost, infrastructure requirements, validation status, and implementation characteristics.

**Results:**
Total of 4,142 records were identified. Thirty sources comprised 17 publications, 3 conference proceedings, and 10 repositories. Six domains of digital pathology solutions were identified: low-cost imaging devices, image viewing and analysis software, AI/deep learning platforms, laboratory information systems, telepathology solutions, and IHC quantification tools. Key innovations included the OpenFlexure microscope (approximately USD 230), DP4ALL scanner-free video-to-whole-slide-image reconstruction, QuPath, Slideflow, DeepLIIF, and OpenELIS. An integrated open-source ecosystem capable of supporting image acquisition, AI-assisted analysis, teleconsultation, biomarker quantification, and laboratory workflow management was identified. Evidence suggests that a functional AI-enabled digital pathology workstation can be assembled for approximately USD 285 using open-source hardware and software.

**Conclusions:**
Low-cost digital pathology technologies provide a feasible pathway for expanding diagnostic, educational, and research capacity in LMICs. However, sustainable implementation requires infrastructure development, workforce training, local validation of AI models, and supportive regulatory frameworks. A phased implementation roadmap is proposed to facilitate scalable adoption in resource-limited settings.

# 174 - Virtual Staining-Based Transfer of Ki-67 Cell Annotations for Scalable H&E Tumor Cell Detection

**Daehong Lee, Kyoung Tai Min**
AIVIS Inc

**Background/Objectives**
Cell-level annotations in hematoxylin and eosin(H&E) histopathology require substantial manual effort, limiting robust tumor cell detection model development. We investigated whether existing Ki-67 immunohistochemistry(IHC) annotations could be repurposed for H&E model training by translating Ki-67 images into synthetic H&E images while preserving cell labels and coordinates.

**Methods**
A Contrastive Unpaired Translation model was trained on 5,000 unpaired patches from breast cancer whole-slide images(WSIs) to generate synthetic H&E images from Ki-67 IHC images. These synthetic images and transferred labels were used to train a Vision Transformer for tumor cell detection and classification. Performance was evaluated on an external Post-NAT-BRCA H&E test set comprising 24 WSIs, 100 regions of interest, 10,665 tumor cells, and 7,659 non-tumor cells. To assess adaptation to the H&E stain domain, the synthetic H&E trained model was compared with a model trained directly on original Ki-67 IHC images using the same labels.

**Results**
On the external test set, the synthetic H&E trained model achieved higher detection recall than the Ki-67 trained model, with values of 0.875 and 0.775, respectively. It also showed higher overall classification accuracy, with values of 0.777 and 0.715, and higher classification macro F1 score, with values of 0.774 and 0.692. Detection precision was low for both models, with values of 0.170 and 0.201, respectively, and should be interpreted with caution because the external test set lacked exhaustive cell annotations. Overall, these results suggest that synthetic H&E training may improve adaptation to the H&E domain compared with direct Ki-67 training.

**Conclusions**
Virtual transfer of Ki-67 annotations to synthetic H&E images may provide a scalable strategy for expanding H&E training data using existing IHC cell labels. Further validation with exhaustively annotated H&E datasets is needed to clarify detection precision and generalizability.

# 176 - 3D False-Color H&E Imaging System for Visualization of Human Tissue Structures

**Sakiko Toyama[1,2], Akira L. Yoshikawa[1,2,3,4], Yuri Saito[1,2], Kohei Otomo[1,2], Junya Fukuoka[2,3,4], Etsuo A. Susaki[1,2]**

[1] Department of Biochemistry and Systems Biomedicine, Juntendo University Graduate School of Medicine
[2] Nakatani Biomedical Spatialomics Hub, Juntendo University Graduate School of Medicine
[3] Department of Pathology, Kameda Medical Center
[4] Department of Pathology Informatics, Nagasaki University Graduate School of Biomedical Sciences

**Background:**
Standard pathological diagnosis primarily relies on two-dimensional (2D) observation of thin sections, which inherently limits the comprehensive assessment of 3D tissue architecture and the continuity of pathological lesions. We have been developing a three-dimensional (3D) false-color hematoxylin and eosin (H&E) imaging system utilizing tissue-clearing technology. This study aims to extend our 3D imaging platform to various human pathological specimens to evaluate its feasibility in visualizing 3D structures for both non-tumor and tumor regions.

**Methods:**
Human pathological specimens (approximately Φ4–6 mm, ~5 mm thick) were utilized for analysis. Tissue clearing was performed using CUBIC technology, followed by 3D fluorescence imaging of nuclei and cytoplasm using a light-sheet microscope (descSPIM). The acquired volumetric data were processed for 3D rendering and converted into false-color H&E-like images using ImageJ to provide visual representations familiar to pathologists.

**Results:**
Despite variations in clearing efficiency depending on specific tissue characteristics, largely uniform staining and transparency were achieved throughout the 5 mm depth across various human organs, enabling successful 3D data acquisition. In non-tumor areas, the system successfully captured vascular networks and the spatial continuity of parenchymal structures. In tumor tissues, particularly in lung adenocarcinoma specimens, we observed the spatial distribution of high-cell-density regions and invasion patterns that are challenging to fully grasp from conventional 2D sections.

**Conclusion:**
The developed system demonstrates significant potential for 3D observation across diverse human organs while maintaining the familiar H&E-like modality. Although technical challenges regarding resolution remain, this approach serves as a valuable 3D complement to traditional 2D histopathology. Ongoing efforts focus on case accumulation and technical optimization to establish 3D-based diagnostic criteria and construct a comprehensive "Human 3D False-Color H&E Atlas". Key words: 3D pathology, stereopathology, tissue clearing, light sheet fluorescence microscopy, False-color H&E Conflict of interest: •Etsuo A. Susaki is an executive officer of CUBICStars, Inc.

# 177 - Automated Nottingham Histologic Grading of Invasive Breast Carcinoma Using a Hierarchical AI Framework on Whole-Slide Images

**IK Zu Quan[1,2], Patsy Ng[3], Md Jahid Hasan[2], Noraidah Masir[4], Shew Fung Wong[1,5], Elaine Wan Ling Chan[6]**

[1] Institute for Research, Development and Innovation, IMU University
[2] Vidanex
[3] Premier Integrated Labs, Corporate Office, Level 21 TNB Dua Sentral
[4] Premier Integrated Labs, Prince Court Medical Centre
[5] School of Medicine, Department of Pathology and Pharmacology, IMU University
[6] Department of Data Science and Artificial Intelligence, Faculty of Engineering and Technology, Sunway University

**Background:**
Nottingham histologic grading is an important prognostic system for invasive breast carcinoma based on tubule formation, nuclear pleomorphism, and mitotic activity. However, manual grading of whole-slide images (WSIs) is labor-intensive and subject to inter-observer variability. This study presents a hierarchical modular artificial intelligence (AI) pipeline for automated Nottingham grading on H&E stained WSIs focused on invasive carcinoma regions.

**Methods:**
A multi-stage AI framework was developed using the BEETLE, BRACS, MIDOG++, and local PIL datasets. An initial "Gatekeeper" model was trained on the BEETLE dataset to identify invasive ductal carcinoma (IDC) regions while excluding benign epithelium, stromal tissue, and ductal carcinoma in situ (DCIS). For tubule formation assessment, invasive regions from BRACS and PIL WSIs were tiled into image patches and analyzed using a weakly supervised learning strategy guided by Nottingham tubule scoring criteria. Patch-level feature embeddings extracted using the UNI foundation model were classified using a lightweight prediction head and aggregated for slide-level scoring. Nuclear pleomorphism assessment employed a similar weakly supervised multi-class framework using invasive patches from Pleomorphism Scores 1–3. Mitotic activity was quantified using a Faster R-CNN object detection model trained on the MIDOG++ dataset to localize mitotic figures within invasive tumor regions. Final Nottingham grades were derived through integration of all grading components.

**Results:**
The Gatekeeper model achieved an F1-score of 0.94 for invasive carcinoma localization. The mitosis detector demonstrated greater than 95% precision and recall on MIDOG++. Tubule formation assessment achieved 100% concordance within the PIL cohort and 93% agreement on the BRACS cohort. Nuclear pleomorphism classification achieved approximately 90% slide-level accuracy. The integrated pipeline achieved overall grading accuracy exceeding 90% within the local validation cohort.

**Conclusion:**
The proposed hierarchical AI framework demonstrated robust performance for automated Nottingham histologic grading across multiple datasets, supporting the potential clinical utility of AI-assisted computational pathology systems.

# 178 - Establishing The first Primary Telepathology and Histopathology in Rural Papua New Guinea: The Kundiawa Experience

**Josephine Jodie Chanoan**
Goroka Provincial Hospital, Eastern Highlands Province

Establishing The first Primary Telepathology and Histopathology in Rural Papua New Guinea: The Kundiawa Experience Lynne Tonar2, Gadelkareem Nassar2, Josephine Chanoan1, Jonathan Itaki1, Ruan Amos1 Joseph Ande3, Kassi Ken3 Pathology Department, Goroka Provincial Hospital, Goroka, Papua New Guinea Pathology Department, Kundiawa General Hospital, Kundiawa, Papua New Guinea Obstetrics & Gynecology Department, Kundiawa General Hospital, Kundiawa, Papua New Guinea Curative & Corporate Department, Kundiawa General Hospital, Kundiawa, Papua New Guinea

**Background**
Histopathology in Papua New Guinea has faced delayed TAT in cancer diagnosis detrimental to patient continuum of care. The Histopathology services have been centralized since colonial days at Port Moresby General Hospital with closure of services in the provinces. The last five years have seen a dramatic shift in decentralization of services with the implementation of the first digital pathology in Kundiawa Hospital, Simbu Province.

**Methods**
A new Histopathology laboratory was built with workflow ergonomics from automated labelling, tissue processing, automated cover-slipping with whole slide imaging (WSI). Images were transmitted with high resolution to the Highland's regional Pathologist.

**Results**
The Histopathology lab set up with first Digital Pathology has seen complex, oncological, non-neoplastic cases with reduced TAT. There has been no specimen referred to the Port Moresby General Hospital.

**Conclusion**
The successful establishment of Provincial Histopathology with telepathology is evident that digital pathology is viable, sustainable and the way forward for PNG. Support in ICT and Pathology networking abroad are crucial for this quality service sustainability. Keywords: Histopathology, Telepathology, Whole Slide Imaging, Kundiawa Hospital

## 179 - Exploring Large Language Model for Cervical Cytology Interpretation

**yayi susanto, Lisnawati Rachmadi, Riesye Arisanty, Marini Stephanie, Eka Susanto**
Departement Of Anatomical Pathology Faculty Of Medicine Universitas Indonesia/ Cipto Mangunkusumo Hospital

**Background/objective:**
Cervical cancer still ranks second most common malignancies in Indonesian women. HPV DNA examination for screening still can't be implemented in many places so that the pap smear is still the chosen screening method. Pap smear examination is a subjective so that the implementation of AI is expected to improve diagnosis accuracy, but currently the use of digital pathology and paid AI algorithms to assist the pap smear diagnosis process is still very limited in Indonesia, however the use of AI in the form of large language models (LLM) is very popular, so in this study we will explore the accuracy of LLM in interpreting pap smear images.

**Methods:**
Ninety eight representative cervical cytology images from Cipto mangukusumo hospital archive and Bethesda system for reporting cervical cytology 3rd edition, were uploaded into GPT 5 – mini and prompted "help me analyzed this cervical pap smear image and provide a diagnosis based on the third edition of Bethesda system for reporting cervical cytopathology". Concordance in each group of diagnosis were calculated.

**Result:**
The ability of GPT 5 – mini to detect abnormalities were 64.7% (44/68), but when grouped into general group of benign, equivocal, and malignant the concordance decreased into 53.33% (16/30), 23,52% (8/34), and 0% (0/34). The concordance according to Bethesda group for NILM 53.33% (16/30), LSIL 80% (8/10), HSIL 44.4% (4/9), while there were 0% concordance for ASCUS, ASCH, SCC, AGC, and adenocarcinoma.

**Conclusion:**
There was fair result of GPT 5 - mini to detect pap smear abnormalities, with the highest concordance for LSIL, but for other Bethesda category the concordance decreased significantly. So that GPT 5 mini not yet able to became a reliable tools for interpreting pap smear result independently.

# 181 - Characterizing Cancer-Associated Fibroblast Subtypes within the Pancreatic Ductal Adenocarcinoma Microenvironment Using Deep Learning and Spatial Proteomics

**Hai Nguyen[1], Rachel Fincham[2], Hemant Kocher[3,4], Joe Yeong[2,5], David Ho[6]**

[1] National Cancer Center, South Korea
[2] Department of Anatomical Pathology, Singapore General Hospital
[3] Centre for Tumour Biology, Barts Cancer Institute-a Cancer Research United Kingdom (CRUK) Centre of Excellence, Queen Mary University of London
[4] Barts and the London Hepato-Pancreato-Biliary (HPB) Centre, The Royal London Hospital, Barts Health National Health Service Trust
[5] Institute of Molecular and Cell Biology (IMCB), Agency for Science, Technology and Research (A*STAR)
[6] Department of Electrical and Computer Engineering, The State University of New York

**Background**

Cancer-associated fibroblasts (CAFs) are dominant orchestrators of the pancreatic ductal adenocarcinoma (PDAC) tumor microenvironment (TME) and functionally heterogeneous. The pCAFassigner classification system is able to link the presence of specific CAF subtypes to clinical outcomes. However, CAF subtyping still requires costly spatial proteomics or transcriptomics assays unavailable in routine pathology. An urgent need exists for a spatially resolved, label-free method of CAF subtype detection applicable to standard hematoxylin and eosin (H&E) whole slide images (WSIs) with the aid of a deep learning technique. Materials and Methods A cohort of 128 PDAC patients with 276 paired TME spatial proteomics images and matched same-section H&E tumor microarrays (TMAs) was assembled. Per-cell biomarker expression profiles were extracted and thresholded to generate curated CAF subtype labels as the ground truth. Labeling identified at least four transcriptomically specific subtypes: A (Periostin), B (Myosin-11), C (Podoplanin), and CD105 lineage (Endoglin), each with unique prognostic implications and immune crosstalk profiles. Spatial proteomic and morphological features were extracted using corresponding foundation models. A deep learning-based pipeline was trained on spatially registered patches to predict CAF subtypes directly from morphology, without reliance on molecular staining. Model was evaluated by 5-fold cross-validation with an area-under-the-curve (AUC) metric.

**Results**

The model achieved AUC performance 0.78 on train data and 0.77 on held-out test data for subtype-level classification, capturing morphological correlates of CAF populations. Using spatially resolved ground-truth, our model provided CAF subtype mapping while requiring only H&E images at inference.

**Conclusions**

This study presents a deep learning-based framework that decodes CAF subtype heterogeneity from routine H&E images by leveraging spatial proteomics. The approach bridges high-plex molecular profiling and scalable digital pathology, offering a clinically translatable tool for CAF-based prognostication and patient stratification in PDAC, as well as a flexible deep learning-based pipeline extensible to other carcinoma types.

# 182 - Enhancing Patient Understanding of Pathology Through Digital Pathology: A Patient Advocacy Initiative by the Japanese Society of Digital Pathology

**Yuka Kitamura[1], Ichiro Maeda[2], Akira Ono[3], Junya Fukuoka[1]**
[1] Nagasaki University
[2] Kitasato University
[3] Shizuoka Cancer Center

**Background:**
Digital pathology and AI are increasingly used in diagnostics, but their potential to improve patient understanding and shared decision-making remains underexplored. Digitized pathology images could serve as visual communication tools, yet opportunities for patient engagement with such images are limited. The role of digital pathology as a tool for patient education, advocacy, and shared decision-making therefore warrants further exploration.

**Methods:**
As a patient advocacy initiative by the Japanese Society of Digital Pathology (JSDP), public educational seminars were held to promote understanding of pathology, digital pathology, and AI. Seminars included sharing whole slide images (WSIs) via a web-based platform, allowing attendees to view WSIs on their own smartphones, followed by discussions. Post-seminar questionnaires assessed participants' views on using pathology images in clinical communication.

**Results:**
Four seminars were conducted (10–20 participants each). Among non-healthcare participants, all agreed that pathology images should be used during outpatient consultations to explain conditions, and 93% agreed such images could help reduce disease- and treatment-related anxiety. Additionally, all attendees favored outpatient services being conducted by pathologists rather than clinicians. Responses indicated that visualizing pathological findings enhances disease comprehension, transparency, and patient-clinician communication.

**Conclusions:**
This JSDP initiative suggests digital pathology can contribute not only to diagnostics but also to patient education, empowerment, and shared decision-making. By making pathological information more accessible, digital pathology supports patient-centered care. This early-phase initiative may serve as a model for patient and public engagement in digital pathology in Japan and the broader Asian region.

# 184 - Automated Nuclear Morphometry of Uterine Smooth Muscle Tumors: A Pilot Study of Usual Leiomyoma, FH-Deficient Leiomyoma, and Leiomyosarcoma

**Aleksandra Asaturova[1], Aleksandra Rogozhina[1], Dmitry Yakovlev[2], Alina Badlaeva[1]**

[1] FSBI «National Medical Research Center For Obstetrics, Gynecology And Perinatology Named After Academician V.I.Kulakov»

[2] MIREA-Russian Technological University

**Background:**

Quantitative nuclear morphometry may provide reproducible features for distinguishing uterine smooth muscle tumors with overlapping histologic appearances. We evaluated whether automated nuclear segmentation and feature extraction could identify group-specific profiles in usual leiomyoma (ULM), fumarate hydratase–deficient leiomyoma (dFH-LM), and leiomyosarcoma (LMS).

**Methods:**

During model development, a fine-tuned Cellpose-CNN achieved the best performance on a training dataset containing 8,825 expert-annotated nuclei, with Dice and panoptic quality scores of 0.8723 and 0.7630, respectively, and was selected for subsequent instance segmentation. Morphometric analysis included 210 hematoxylin-and-eosin image patches from uterine nodules of 21 patients, with seven patients per group and 10 patches per patient. The model segmented 6,976 nuclei: 2,309 in ULM, 1,758 in dFH-LM, and 2,909 in LMS. Nuclear size, shape, staining intensity, hyperchromasia, and spatial density were quantified. Patient-level values were summarized as medians and interquartile ranges. Groups were compared using the Kruskal–Wallis test, followed by Mann–Whitney U tests with Holm adjustment.

**Results:**

Mean hematoxylin intensity differed significantly among groups (P=0.039), with median values of 0.065 in ULM, 0.091 in dFH-LM, and 0.084 in LMS. The fraction of hyperchromatic nuclei also differed significantly (P=0.033): 3.5%, 16.1%, and 9.4%, respectively. The ULM–LMS difference remained significant after Holm correction (adjusted P=0.033). dFH-LM showed the lowest nuclear count and spatial density, whereas LMS had the largest nuclear area. ULM nuclei were smaller and less intensely stained but had the highest eccentricity, consistent with a more elongated nuclear shape.

**Conclusion:**

Automated instance segmentation identified distinct morphometric patterns across ULM, dFH-LM, and LMS. Nuclear chromatin intensity and hyperchromasia showed the strongest group-level differences and may support computational classification of uterine smooth muscle tumors. These exploratory findings require validation in larger independent cohorts.

# 188 - Improving Segment Anything with a Decoder-Attached Classification Head for Automated Diagnosis of Liquid-Based Cervical Cytology Smears

**Aleksandra Asaturova[1], Anna Timakova[2], Anna Tregubova[1], Arsenii Litvinov[3], Eugeny Karpulevich[3], Aleksandra Rogozhina[1], Alena Shadrina[1]**

[1] FSBI «National Medical Research Center For Obstetrics, Gynecology And Perinatology Named After Academician V.I.Kulakov»

[2] Sechenov University

[3] Trusted AI Research Center RAS

**Background:**

Cervical cancer screening via liquid-based cytology remains a cornerstone of prevention, yet manual review of whole-slide images is labor-intensive and prone to inconsistency. Foundation models like SAM excel at segmentation but lack native classification abilities tailored to cytological abnormality grading. We investigated whether a lightweight modification - a classification head appended to SAM's decoder - could enable joint cell detection and multi-class diagnosis directly from digitized smears.

**Methods:**

We utilized a dataset of 2900 cells from liquid-based cervical cytology whole slide images from the Kulakov National Medical Research Center. Diagnoses comprised NILM, HSIL, LSIL, ASC-US and ASC-H. The images were splitted into training and validation sets. We extended a pre-trained SAM backbone by attaching a task-specific classification head to its decoder, allowing the model to simultaneously propose cell regions and predict 5 classes. Training was conducted with standard augmentation and validated on the held-out evaluation subset. Metrics included detection-oriented metrics (mAP@[0.5,0.75], NoC@0.8, NoC@0.9) and classification metrics (AUROC, F1-score, accuracy, specificity, precision, recall).

**Results:**

These methods allowed us to achieve balanced detection and classification metrics: AUROC=0.88, F1-score=0.49, Precision=0.50, Recall=0.49, NoC @0.8=3.50, NoC@0.9=5.07, mAP@0.5=0.13, mAP@0.75=0.09.

**Conclusion:**

Adding a dedicated classification head to SAM's decoder represents a feasible, minimally invasive strategy for automated cervical cytology interpretation. The encouraging results warrant scale-up studies, inclusion of more diverse centers and pathway integration to support cytologists in routine practice. The funding information: the research was funded by the Russian Science Foundation, grant No. 25-75-20027

# 189 - BRIGHT: A Collaborative Generalist-Specialist Foundation Model for Breast Pathology

**Xiaojing Guo[1], Jiatai Lin[2], Yumian Jia[1], Fangfang Liu[1], Chu Han[2], Zaiyi Liu[2]**

[1] Department of Breast Pathology and Laboratory, Tianjin Medical University Cancer Institute \& Hospital, National Clinical Research Center for Cancer, Key Laboratory of Breast Cancer Prevention and Therapy, Tianjin Medical University, Ministry of Education, Tianjin's Clinical Research Center for Cancer

[2] Guangdong Provincial People's Hospital

**Background:**

Generalist pathology foundation models (PFMs), pretrained on large-scale multi-organ datasets, have demonstrated remarkable predictive capabilities across diverse clinical applications. However, their proficiency across the full spectrum of clinically essential tasks within a specific organ system remains unclear due to the lack of large-scale validation cohorts for a single organ and the absence of a tailored training paradigm that effectively translates broad histomorphological knowledge into organ-specific expertise.

**Objective:**

To develop and validate BRIGHT, the first PFM specifically designed for breast pathology, and to systematically evaluate its performance across a comprehensive set of clinical tasks in breast oncology.

**Methods:**

BRIGHT was trained on over 51,000 breast whole-slide images from over 40,000 patients across 19 hospitals, employing a collaborative generalist-specialist framework. For validation, we curated multi-institutional cohorts comprising over 25,000 WSIs from 10 hospitals, covering 25 distinct clinical tasks spanning diagnosis, biomarker prediction, treatment response, and survival prediction.

**Results:**

BRIGHT outperformed five leading generalist PFMs, achieving state-of-the-art performance in all 25 internal validation tasks and in 4 of 11 external validation tasks, with excellent heatmap interpretability.

**Conclusions:**

This large-scale validation demonstrates BRIGHT's clinical utility in breast oncology and validates a collaborative generalist-specialist paradigm, providing a scalable template for developing organ-specific PFMs to accelerate the translation of foundation models into specialized clinical practice.

# 190 - Unsupervised Artificial Intelligence Clustering of Tumor Architectural Patterns Identifies Reproducible Precursors of Spread Through Air Spaces in Lung Adenocarcinoma

**Mana Saito[1], Yukio Tsushima[2], Kris Lami[1], Ryuta Matsuda[1], Junya Fukuoka[1]**

[1] Department of Pathology Informatics, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan

[2] Shinmatsudo Central General Thoracic Surgery, Chiba, Japan

**Background/Objective:**

Although spread through air spaces (STAS) serves as a critical prognostic indicator in lung adenocarcinoma, its underlying morphological characteristics are not well established. This study sought to discover reproducible morphological signatures of tumor cells correlated with STAS via unsupervised machine learning techniques.

**Methods:**

Two independent cohorts of surgically resected lung adenocarcinomas (n=100 each, total n=200) were evaluated. At the tumor periphery, STAS-positive zones were identified as having five or more STAS clusters within a 2-mm radius. To ensure a balanced representation, STAS-negative regions (up to four per patient) were correspondingly annotated. Circular regions of interest (ROIs) measuring 2 mm were extracted and segmented into 224×224 pixel patches via QuPath. A custom image-sorting tool filtered the data, selecting patches with at least 70% tumor cellularity while discarding those with 75% or more background tissue. The retained patches were then processed through a large Vision Transformer (ViT) to produce 1024-dimensional feature vectors, followed by independent unsupervised clustering (k=50) for each cohort.

**Results:**

Cluster distribution remained stable across the cohorts, with median patch counts of 115 and 90. Six specific clusters exhibited a robust, reproducible correlation with STAS-positive areas (defined as ≥70% prevalence in both datasets), with some demonstrating exceptional specificity (97% to 100%). In contrast, eight clusters were tightly linked to STAS-negative zones (≤30% prevalence). Key STAS-associated clusters revealed distinct morphological phenotypes, such as invasive mucinous adenocarcinoma, mucin-abundant acinar adenocarcinoma, micropapillary patterns within acinar or solid structures, complex/cribriform acinar architecture, and solid adenocarcinoma accompanied by substantial inflammatory infiltration.

**Conclusion:**

The application of unsupervised clustering successfully isolates reproducible, STAS-related morphological phenotypes in tumor cells. Establishing these visual signatures offers a solid groundwork for improved pathological evaluation and future predictive modeling, ultimately facilitating the creation of AI-driven diagnostic tools for STAS in limited biopsy samples.

# 193 - Digital Quantification of Mature Tertiary Lymphoid Structures with Germinal Center (TLS-GC) Enables Risk Stratification in Oral Cancer—A Whole-Slide Analysis of 493 Patients

**Mrudula Katarni[1], Shireen Ali[1], Parvathy Nath[2], Washim Saahil[3], Anshuman Sharma[3], Saurabh Nagar[1], Asawari Patil[4], Neha Mittal[2], Munita Bal[2], Sarbani Ghosh-Laskar[5], Richa Vaish[6], Tripti Bameta[3], Swapnil Rane[1]**

[1] Pathology,DCO,ACTREC,HBNI
[2] Pathology,TMH,HBNI
[3] DCO,ACTREC,HBNI
[4] Pathology,ACTREC,HBNI
[5] Radiation Oncology, TMH, HBNI
[6] Surgery, TMH, HBNI

**Background/Objective**

Mature tertiary lymphoid structures with germinal centres (TLS-GC) are indicators of the body's tumor-specific humoral immune response, but little is known about their prognostic impact. This study evaluated whether enumerating TLS-GC and tertiary lymphoid structures without germinal centres (TLS) on digital whole-slide images could yield reliable immune-related markers to predict clinical outcomes in OSCC patients.

**Method**

We retrospectively reviewed H&E-stained whole-slide images (WSIs) of primary tumors from 493 locally advanced(T3/4 or N2/3) chemo-naive surgically treated OSCC cases. TLS were quantified by blinded observers and classified into those without GC (TLS) and those exhibiting a germinal center (TLS-GC). These counts were summarised at patient level and slide level, followed by clinicopathological analysis and Survival analysis using Kaplan–Meier curves and multivariate COX regression models

**Results**

TLS & TLS-GC were noted beyond the tumor boundaries. TLS were found in 159/493(32.3%) cases and TLS-GC were seen in 128/493(26%) cases. The mean TLS count & TLS-GC counts/slide were 4±0.340 & 1±0.097, respectively. Degree of differentiation was associated with TLS-GC density, with well-differentiated tumors having more TLS-GC than poorly differentiated tumors (p=0.006). Presence of TLS-GC was associated with significantly better OS (HR:0.421,95% CI 0.280–0.634,p<0.0001), DFS (HR:0.536,95% CI 0.390–0.736,p<0.0001), and LRC (HR:0.610,95% CI 0.437–0.850,p=0.003) as compared to those without TLS or those with TLS. There was no association between TLS-GC counts with GC counts in regional nodes, s-TIL, TNM stage, gender, age, adjuvant therapy arms and other clinicopathological factors. Multivariate analysis revealed that germinal center maturation is an independent predictor of short-term and long-term outcomes.

**Conclusion**

Our study highlights critical insights by demonstrating germinal center maturation within TLS, rather than bulk TLS density, predicts patient outcomes. These findings support that quantitative evaluation of TLS maturation could add to current prognostic models. Our ongoing work focuses on training a deep-learning model to automate this immune quantification workflow.

# 194 - Building Consensus for Pixel-Wise Annotation of Pathology Whole Slide Images: Experience and Lessons from a Multi-national Collaborative Study

**Hang Nguyen[1], Hai Nguyen[2], Maximus Yeung[3], Horace Cheung[3], Mohamed Yakoub[4], Narasimhan Agaram[4], Meera Hameed[4], Poonam Panjwani[5], Bharat Rekhi[5], Mukta Ramadwar[5], Truong Nguyen[6], Jijgee Munkhdelger[7], Mirzabek Shodmonov[8], Gulrukh Botiralieva[8], Le Phi Nguyen[9], Andrey Bychkov[7], David Ho[10]**

[1] Hanoi Medical University
[2] National Cancer Center
[3] The University of Hong Kong
[4] Memorial Sloan Kettering Cancer Center
[5] Tata Memorial Hospital, Homi Bhabha National Institute (HBNI) University
[6] University of Medicine and Pharmacy at Ho Chi Minh City
[7] Kameda Medical Center
[8] Ipsum Pathology
[9] Institute for AI Innovation and Societal Impact (AI4LIFE), Hanoi University of Science and Technology
[10] The State University of New York, Korea

**Background/Objective:**
Accurate pixel-wise annotations are essential for training deep learning-based segmentation models for whole slide images (WSI). Substantial inter-observer variability among annotators can affect annotation quality and hinder model development. To address this challenge, consensus-building among experts from multiple collaborating institutions is increasingly important. Histopathological assessment of post-treated osteosarcoma specimens is complex, challenging and time-consuming, but clinically important. In this study, we describe the preparation, implementation, and lessons learned from a multi-national consensus aimed at harmonizing osteosarcoma annotations, including neo-adjuvant treated scenarios.

**Methods:**
The study consortium included 13 pathologists and 2 AI developers from seven countries. Twelve challenging scenarios representing common disagreements in post-treatment osteosarcoma pathology were distributed two weeks before the meetings. The scenarios covered hemorrhage, fibrinoid change, cystic degeneration, osteoid-rich areas, treatment-related changes, chondroblastic differentiation, soft tissue invasion, and reactive bone/cartilage. To accommodate multiple time zones, two online consensus meetings were scheduled. During the online consensus meetings, WSIs were reviewed interactively. Discussions were recorded for annotators unable to attend, and consensus statements were documented. One unresolved scenario was subsequently circulated through an online survey to establish final consensus.

**Results:**
Five pathologists attended the first consensus meeting and 6 pathologists attended the second meeting. The two AI developers also participated in the meetings to provide technical input on annotation strategies and model development. Eleven of 12 annotation scenarios reached consensus during the meetings. The one remaining scenario involving geographic intratumoral hemorrhage with admixed red blood cells and infarct-like necrosis was resolved through a follow-up survey.

**Conclusion:**
A structured multi-national consensus building is feasible and effective for harmonizing pixel-wise pathology annotations in rare tumor studies. Early distribution of representative scenarios, interactive WSI review, consensus documentation, and post-meeting surveys were key factors in achieving agreement. These experiences may serve as a framework for future international computational pathology collaborations.

# 196 - Deep Learning-Based Distinction of Dedifferentiated Liposarcoma and Leiomyosarcoma on Histopathology Reveals Challenges in Morphologically Ambiguous Cases

**Wanyi Lin[1], Martijn P.A. Starmans[1,2], Stefan Klein[1], Geert J.L.H. van Leenders[2], Farhan Akram[2], Karthik Prathaban[1]**

[1] Department of Radiology and Nuclear Medicine, Erasmus MC Cancer Institute, University Medical Center

[2] Department of Pathology and Clinical Bioinformatics, Erasmus MC Cancer Institute, University Medical Center

Background/Objective : Dedifferentiated liposarcoma (DDLPS) can exhibit spindle-cell morphology that resembles leiomyosarcoma (LMS), making their distinction on routine histopathology challenging in some cases. Consequently, ancillary molecular testing, including MDM2 amplification analysis, is often required for diagnostic confirmation. While AI-based digital pathology has shown promise in common tumors, its application in soft tissue sarcomas remains limited. We investigated whether a deep learning model could be developed to distinguish DDLPS from LMS on hematoxylin and eosin-stained (H&E) whole-slide images (WSIs).

**Methods:**
H&E-stained WSIs from 56 DDLPS patients (85 slides) and 96 LMS patients (151 slides) from the public TCGA-SARC dataset were used for model training and internal evaluation. A deep learning model was developed to generate patient-level predictions directly from whole-slide images without manual tumor annotations. External validation was performed using an independent local cohort comprising 10 DDLPS and 10 LMS.

**Results:**
The model achieved an area under the curve (AUC) of 0.95 on the held-out TCGA test cohort and 0.90 in external validation. All external DDLPS cases were correctly classified, whereas 4 of 10 LMS cases were misclassified. Review of pathology reports identified five LMS cases that required MDM2 amplification testing during routine diagnostic evaluation. Of these diagnostically challenging cases, four were misclassified by the model as DDLPS, while the fifth was correctly classified but demonstrated near-equivalent class probabilities (0.53 LMS vs 0.47 DDLPS).

**Conclusion:**
These findings demonstrate the feasibility of deep learning-based WSI analysis for distinguishing DDLPS from LMS and suggest that model uncertainty may preferentially arise in diagnostically ambiguous cases requiring ancillary molecular confirmation. These observations support further investigation of AI-based computational approaches for supporting soft tissue tumor diagnosis.

# 198 - Never-smoker under-representation in training data reduces AI performance for EGFR mutation prediction in Singapore lung adenocarcinoma patients

**Corinna Jie Hui Goh[1], Tianyi Zhang[1], Yiling Liu[1], Chi Phan[1], Dasun Premathilaka[1], Shangqing Lyu[1], Mengyuan Pang[2], Jia Chi Yeo[2], Ngak Leng Sim[2], Lanying Wang[3], Angela Takano[4], Anne James[4], Hlaing Nwe Oo[3], Aisha Peng[1], Kok Haur Ong[1], Seow Ye Heng[4], Meixin Shen[3], Xin Xiu Sam[4], Lee Hong Song[4], Chee Leong Cheng[4], Gillianne Geet Yi Lai[3], Stephanie Pei Li Saw[3], Daniel Shao Weng Tan[3], Anders Jacobson Skanderup[2], Tony Kiat Hon Lim[4], Hwee Kuan Lee[1,5,6,7,8], Malay Singh[1]**

[1] Bioinformatics Institute, A*STAR
[2] Genome Institute of Singapore, A*STAR
[3] National Cancer Centre Singapore
[4] Department of Pathology, Singapore General Hospital
[5] School of Computing, National University of Singapore
[6] School of Biological Sciences,Nanyang Technological University
[7] International Research Laboratory on Artificial Intelligence, Institute for Infocomm Research, A*STAR
[8] Centre for Frontier AI Research (CFAR), A*STAR

**Background**

Lung cancer in never-smokers (LCINS) is a growing disease, particularly in Asia. Epidermal growth factor receptor (EGFR) mutations are more prevalent in never-smokers and guide targeted lung cancer treatment decisions. Artificial intelligence (AI) models applied to histopathology whole-slide images could support early triage for EGFR molecular testing. However, lung adenocarcinoma in never-smokers and smokers may follow distinct histologic and molecular pathways. EGFR-driven biology is more frequent in never-smokers while smoking has been associated with solid histologic components enrichment in lung adenocarcinoma. We investigated whether smoking-status imbalance in training datasets influences WSI-based AI performance, particularly for EGFR mutation prediction.

**Methods**

We analysed samples from a Singapore cohort (N=265) and The Cancer Genome Atlas lung adenocarcinoma cohort (N=513). We evaluated the effect of varying the proportion of never-smokers in training datasets on multi-task AI model performance for predicting EGFR mutation status, staging, and relapse status. Performance trends were assessed using weighted F1 scores.

**Results**

Alignment of smoking-status composition between training and test sets improved model performance, particularly for EGFR mutation prediction. In never-smoker-dominant test sets, performance improved as never-smoker representation in the training increased. In smoker-dominant test sets, performance improved with greater smoker representation in training. On never-smoker test samples, models trained on smoker-only samples achieved mean weighted F1 score of 0.363 (SE 0.018) across models and folds for EGFR mutation prediction, while models trained on never-smoker-only samples achieved 0.620 (SE 0.019). Smoking-status composition had minimal effect on predictive performance for stage and relapse status.

**Conclusion**

Smoking-status misalignment between training and deployment populations can reduce robustness and generalisability of lung cancer AI models, particularly for EGFR mutation prediction. As smoking prevalence declines globally, enriching training datasets with never-smoker cases will be important for improving model reliability in majority never-smoker populations and for mitigating bias related to dataset composition.

# 199 - From Narratives to Knowledge Graphs: LLM-Based Extraction and Ontology Alignment of Soft Tissue Tumour Pathology Reports

**Karthik Prathaban[1,2], Farhan Akram[2], Stefan Klein[1], Martijn P.A. Starmans[1,2]**

[1] Department of Radiology and Nuclear Medicine, Erasmus MC Cancer Institute, University Medical Center

[2] Department of Pathology and Clinical Bioinformatics, Erasmus MC Cancer Institute, University Medical Center

**Background/Objective:**

Pathology reports contain rich diagnostic terminology and morphologic detail essential for clinical interpretation, yet their free-text format limits large-scale computational reuse and makes manual extraction laborious. Large language models (LLMs) show potential in extracting structured information from pathology narratives, but variability in diagnostic phrasing and compositional terminology poses challenges for standardization and cross-report analysis. This study integrates an LLM-based extraction pipeline with ontology-guided normalization of extracted tumour diagnoses to support knowledge-graph-based integration of pathology findings across reports.

**Methods:**

We applied an open-weight LLM (DeepSeek-R1-0528-Qwen3-8B) to 50 hand-curated soft tissue tumour reports to extract specimen-level findings, including diagnoses, macroscopic findings, histologic findings, and molecular marker results. The LLM was instructed to represent the results in a knowledge-graph-compatible format (e.g. “specimen_1::shows_tumor_diagnosis::lipoma”, “specimen_2::shows_marker_result::MDM2::negative”). Unique diagnosis terms were normalized against the National Cancer Institute Thesaurus (NCIt), first using exact synonym matching. When no exact match was available, the most semantically related ontology concept was identified automatically. Extraction outputs and ontology mappings were evaluated manually.

**Results:**

The extraction pipeline achieved a macro-average F1-score of 0.91 while capturing clinically relevant relationships between morphologic findings and tumour diagnoses (e.g., “shows_histologic_feature::mitosis::present” for pleomorphic liposarcoma). Of 34 unique extracted tumour diagnosis terms, 20 were matched to NCIt concepts through exact synonym matching. For two additional diagnoses, semantic retrieval identified the correct NCIt concept and corrected a translation error. Of the remaining 12 unmatched diagnoses, 10 were judged partial semantic matches, typically preserving the underlying tumour diagnosis while losing descriptive qualifiers (e.g., “myxofibrosarcoma-like dedifferentiated liposarcoma” to “Dedifferentiated Liposarcoma”). Two diagnoses were judged mismatches (e.g., “low-grade liposarcoma” to “Low-Grade Sarcoma”).

**Conclusion:**

LLM-based extraction pipelines can generate structured pathology representations from free-text reports, while ontology-guided normalization improves standardization and interoperability across heterogeneous diagnostic terminology. These findings enable integration of pathology findings across reports for large-scale analyses of tumour characteristics and outcomes.

# 200 - Artificial Intelligence–Assisted Screening in Thyroid Fine-Needle Aspiration Cytology: Comparative Evaluation of Domain-Specific AI Models, General-Purpose Large Language Model, and Cytopathologist Interpretation

**Mimna V. M[1], Meena Singh[1], Mithlesh Bhargav[1], Jasleen Kaur Oberoi[1], Pankuri Mittal[2]**

[1] Department of Pathology, All India Institute Of Medical Sciences (AIIMS), Gorakhpur, Up, India

[2] Department of ENT, All India Institute of Medical Sciences (AIIMS), Gorakhpur

Multimodal large language models (LLMs) are increasingly applied to cytology microphotographs, but their performance relative to purpose-built classifiers remains poorly characterized. We compared a domain-specific AI model, a general-purpose multimodal LLM, and cytopathologist interpretation (unaided and domain-specific AI-assisted) for binary risk stratification of thyroid FNAC images. A domain-specific image-classification model was developed using 319 MGG-stained thyroid FNAC images from 80 cases (40 benign, 40 higher-risk). External validation used 52 independent cases (24 benign, 28 higher-risk) against a predefined reference standard (histopathology or two-cytopathologist consensus). Four representative MGG images per case were independently evaluated by (1) a cytopathologist, (2) the AI model, (3) a cytopathologist assisted by the AI model, and (4) a multimodal LLM (ChatGPT 5.5) applied zero-shot with a standardized image-only prompt; the LLM additionally assigned a Bethesda category. Diagnostic performance was assessed using accuracy, sensitivity, specificity, and Cohen's κ; paired modalities were compared with McNemar tests, and Bethesda-category concordance with quadratic-weighted κ. The AI-assisted cytopathologist achieved the best performance (accuracy/sensitivity/specificity: 92.3/92.9/91.7%), followed by the unaided cytopathologist (88.5/85.7/91.7%) and the AI model (82.7/89.3/75.0%). The LLM matched the highest sensitivity (92.9%) but had the lowest specificity (50.0%) and accuracy (73.1%; κ=0.44), misclassifying 12/24 benign cases while missing only 2/28 higher-risk cases; specificity was significantly lower than both cytopathologist reads (p=0.006). Agreement between the AI model and cytopathologist interpretation was substantial to almost perfect (κ=0.73-0.92), whereas the LLM agreed only fairly with other modalities (κ=0.31-0.37). Bethesda concordance was poor (exact agreement 30.8%; weighted κ=0.35), with systematic over-grading and no indeterminate-category calls (0/15 Bethesda III cases). A domain-specific AI model and AI-assisted cytopathologist outperformed a general-purpose multimodal LLM, which was highly sensitive but markedly non-specific and over-graded benign aspirates. Current zero-shot LLMs appear unsuitable for autonomous thyroid FNAC triage, though their high negative predictive value suggests a rule-out role pending specificity-focused calibration.

# 202 - Artificial Intelligence-Based Tumor Stroma Ratio as a Prognostic Marker in Oral Squamous Cell Carcinoma: A Retrospective Observational Study.

**Rajasa Jialdasani[1], Sujit Joshi[2], Udita Dewan[3], Abhijeet Patil[4], Shanu Pratap[4], Gourav Agrawal[5], Sharvari Vishwasrao[6], Nupur Mankad[1], Sangeeta Desai[7]**

[1] SenzanAI
[2] Excel Diagnostics
[3] Dewan Histopathology Center
[4] Pixonate Labs
[5] Jupiter Hospital
[6] Independant Consultant
[7] Former Prof and Head Pathology, Tata Memorial Center

**Background:**
Tumor stroma ratio (TSR) is an emerging histomorphologic biomarker in oral squamous cell carcinoma (OSCC), with stroma-rich tumors generally associated with adverse outcomes. This study evaluated whether AI-derived TSR correlates with established adverse clinicopathologic parameters in oral cancer and whether it may help refine prognostic assessment.

**Methods:**
Fifty Eight resected OSCC cases from 2 centers were analysed retrosepctively. Representative regions from the invasive tumor front were selected from digitized H&E WSIs and assessed using a deep learning–based semantic segmentation model developed for automated quantification of tumor and stromal compartments. AI-derived TSR values were correlated with histopathological parameters including tumor size, depth of invasion (DOI), lymph node metastasis (LNM), etc. A TSR cutoff of 50 was considered clinically relevant, with values of 50 or above indicating stroma-rich tumors. The dataset was assessed for trends linking AI-TSR to aggressiveness-related parameters. AI-TSR visual overlays were qualitatively assessed by the pathologists.

**Results:**
Higher AI-TSR values were observed in 35/58 (60.3%) of all tumors, 21/29 (72.41%) with deep invasion, 14/22 (63.64%) with LNM. Of the 27 early stage, pT1-T2 tumors, 6/15 (40%) with high AI-TSR had LNM, while only 2/12 with low AI-TSR (16.6%) had LNM. Among the well differentiated tumors, 15/21 had high AI-TSR and among these 4 had LNM and 5 had DOI more than 10mm. Although statistical significance was limited by sample size and case heterogeneity, the biological trend supports AI-TSR as a potential prognostic surrogate especially in early stage OSCC.

**Conclusion:**
AI-TSR mirrors tumor aggressiveness and shows an unfavorable association with increased DOI, and LNM. Identification of stroma-rich tumors using an AI-based approach may provide a practical adjunct to routine histopathologic risk stratification in early stage OSCC. Larger studies with outcome correlation are needed to validate AI-TSR as an independent prognostic biomarker in OSCC.

# 204 - Clinical Implementation and Validation of a Digital Pathology Remote Telepathology Frozen Section Reporting Service in NUHS

**Yee Shuen, Darren Chua[1], Jaslyn Jie Lin Lee[1], Jeffrey Lum[1], Char Loo Tan[2,1], Gareth David Huw Turner[1]**

[1] Department of Pathology, National University Hospital

[2] Parkway Laboratory Services Ltd

**Background**

To support a centralised hub-and-spoke model for delivering pathology services across the National University Health System (NUHS) cluster, optimising manpower and laboratory space resources, a digital pathology system was adopted as a solution to deliver intraoperative frozen section reports via remote telepathology reporting, removing the previous requirement for an on-site pathologist. We describe our experience with the commissioning, implementation, and validation of a whole slide imaging Digital Pathology system to support frozen section services at Ng Teng Fong General Hospital (NTFGH).

**Methods**

We analysed a pre-implementation validation exercise for all reporting pathologists, and benchmarked accuracy (including any diagnostic discrepancies in reporting) and turnaround times for frozen sections over a one-year period after implementation of digital reporting, compared to the one-year period prior to implementation with traditional glass slide reporting.

**Results**

In the validation exercise, pathologists had similar diagnostic accuracy using digital pathology compared to glass slides, with major discrepancy rates <5% and intra-observer concordance rates ≥95%. The time taken to reach a diagnosis during the validation exercise was significantly greater for digital images than glass slides. However, no significant change in turnaround time or accuracy was seen after implementation of digital reporting despite an increase in frozen section workload.

**Conclusion**

The use of digital pathology for remote telepathology reporting of frozen sections at NTFGH is successful, both in terms of measurable outcomes and implementation.

# 206 - Beyond Image Quality: Assessing Workflow Efficiency, Storage Burden, and Focus Performance Across Specimen Complexity in Whole-Slide Imaging

**Sampada Gursale[1,2], Rajiv Kumar Kaushal[1,2], Subhash Yadav[1,2], Tanuja Shet[1,2]**
[1] Tata Memorial Centre
[2] Homi Bhabha National Institute

**Background:**
As whole-slide imaging (WSI) becomes integral to routine pathology practice, optimizing workflow efficiency while maintaining image quality is increasingly important. The influence of specimen complexity on operational performance and quality-control metrics remains less well characterized. This study assessed WSI performance indicators across biopsy, bone marrow, and resection specimens using two digital pathology platforms.

**Methods:**
A total of 141 routine pathology cases comprising biopsies (n=75), bone marrow specimens (n=26), and resections (n=40) were digitized using two high-throughput WSI systems on auto-mode. Performance metrics included scan acquisition time, file size, scan failures, focus accuracy, image quality, stitch artifacts, and pre-analytical slide-related artifacts. Digital slides were reviewed and comparative analyses were evaluated for workflow efficiency, storage requirements, and image-quality metrics across specimen types and scanner platforms.

**Results:**
Specimen complexity influenced all measured workflow parameters. Scan acquisition times were consistently lower on scanner A than scanner B for biopsies (81 vs. 99 s/slide), bone marrow specimens (85 vs. 157 s/slide), and resections (109 vs. 150 s/slide). Corresponding file sizes (GB/slide) were 0.57 vs. 1.10, 0.52 vs. 1.71, and 1.39 vs. 2.68, respectively. Across the entire cohort, scanner A reduced total scanner occupancy time by approximately 30% and generated 52% less digital storage than scanner B. The greatest differences in acquisition time and storage burden were observed in bone marrow specimens. A throughput–quality trade-off was evident, with scanner B demonstrating fewer out-of-focus regions across biopsies (17.3% vs. 28.0%), bone marrow specimens (19.2% vs. 26.9%), and resections (5.0% vs. 15.0%). Stitch artifacts and scan failures were uncommon, while pre-analytical artifacts had minimal impact on diagnostic interpretation. All digital slides were diagnostically interpretable.

**Conclusions:**
Workflow efficiency, storage utilization, and focus performance represent complementary quality-control metrics that should be considered during digital pathology implementation. This multidimensional assessment provides practical framework for optimizing high-throughput WSI workflows.

# 207 - Does Decalcification Challenge Whole-Slide Imaging ? A Comparative Evaluation of Two High-Throughput Digital Pathology Scanners for Decalcified Tissue and Bone Marrow Specimens

**Sonali Dalvi[1,2], Rajiv Kumar Kaushal[1,2], Subhash Yadav[1,2], Tanuja Shet[1,2]**
[1] Tata Memorial Centre
[2] Homi Bhabha National Institution

**Background:**
Whole-slide imaging (WSI) is increasingly used in routine diagnostic practice. However, decalcified tissues and bone marrow biopsies remain challenging specimens due to their tissue complexity and focal-plane variability. This study compared the performance of two high-throughput WSI platforms for digitizing these specimen types.

**Methods:**
A total of 146 slides, comprising 86 decalcified tissue and 60 bone marrow specimens, were digitized using two high-throughput WSI systems (A and B) using auto-mode. Performance metrics included scanning time, file size, scan failures, image quality, and frequency of out-of-focus (OOF) regions. Digital slides were evaluated for diagnostic adequacy, out-of-focus regions, scan failures, stitch artifacts, and pre-analytical slide-related artifacts.

**Results:**
In the decalcified cohort (n=86), scanner A demonstrated significantly shorter mean scanning times than scanner B (114.8 vs. 139.0 seconds/slide; $p<0.001$) and generated substantially smaller files (1.10 vs. 2.03 GB/slide; $p<0.00$, %, 46% reduction). OOF regions were identified in 24/86 (27.9%) and 30/86 (34.9%) slides scanned by scanners A and B, respectively. No scan failures occurred on either platform. In the bone marrow cohort (n=60), scanner A again achieved faster image acquisition (82.9 vs. 132.8 seconds/slide; $p<0.001$) and smaller file sizes (0.61 vs. 1.46 GB/slide; $p<0.001$, 58% reduction). Scanner B demonstrated superior focus performance, with OOF regions observed in 9/60 (15.0%) slides compared with 21/60 (35.0%) on scanner A. One scan failure occurred on scanner A, whereas none were observed on scanner B. Stitch artifacts were rare on both platforms. Pre-analytical artifacts, predominantly air bubbles and pen markings, were infrequent and did not affect diagnostic interpretation. Both systems generated diagnostically acceptable images with minimal artifacts.

**Conclusions:**
Both WSI platforms were suitable for routine diagnostic use. Scanner A provided superior throughput and storage efficiency, whereas focus performance varied by specimen type, underscoring the need for specimen-specific validation during digital pathology implementation.

# 209 - Spatial Gradient-Based Zonal Modeling Reveals Prognostically Relevant Tumor Microenvironment Patterns in Breast Cancer

**Zhongze Cui, Meng Yue, Si Wu, Jiaxian Miao, Yueping Liu, Yan Ding**
Department Of Pathology, The Fourth Hospital Of Hebei Medical University

**Background:**
The tumor microenvironment (TME) is closely associated with breast cancer prognosis, yet conventional assessments lack refined spatial zoning and cannot determine the optimal distance from tumors for prognostically valuable information.

**Methods:**
This study included 1,190 breast cancer patients (1,384 WSIs) from TCGA-BRCA, CPTAC, and three external cohorts. We first developed Usegpathformer, a pixel-level tumor segmentation model integrating UNI and SegFormer. Using a boundary dilation-erosion strategy anchored to tumor nest borders, we divided the TME into 10 functional zones: pure tumor, intratumoral stroma, tumor region, three outward expansions from the tumor region (0.1, 0.5, 1.0 mm), three outward expansions from the intratumoral stroma (0.1, 0.5, 1.0 mm, excluding tumor), and the whole tissue. Deep features were extracted from each zone to evaluate prognostic value for overall survival, with mechanistic exploration via heatmaps and multiplex immunofluorescence.

**Results:**
Usegpathformer achieved robust segmentation performance (Dice: 0.80–0.87 across all cohorts). Prognostic analysis on TCGA showed that in the validation set, the C-index for pure tumor was 0.6747, while for whole slide images it was 0.6896. As stromal regions were incorporated, the C-index exhibited an "inverted U-shaped" trend, peaking at 0.1 mm (0.7115) and 0.5 mm (0.7110) expansion zones. For pure stromal regions, the C-index rose rapidly and stabilized at 0.5 mm. These trends were validated in external cohorts. Heatmaps revealed that the model focused on disorganized (poor prognosis) rather than lymphocyte-rich or homogeneous (good prognosis) stroma. Multiplex immunofluorescence showed that low-risk patients had significantly higher densities of $CD3^+$, $PD1^+$, and $CD3^+/PD1^+$ T cells.

**Conclusions:**
This study identifies that the stroma within 0.5 mm of tumor expansion represents both the peak prognostic region (for tumor-plus-stroma combined zones) and the stabilization threshold for pure stromal prediction. This framework provides a quantifiable TME assessment tool for pathologists and offers insights for precise immune microenvironment prediction.

# 210 - Comparison of morphological diagnosis and semi-quantitative findings of native renal biopsy on routine light microscopy with digital whole slide imaging.

**Manoj Jain**
Department of Pathology, SGPGIMS, Lucknow

**AIM:**
To compare and assess semi-quantitative findings and morphological diagnosis of native renal biopsy on routine light microscopy with digital whole slide imaging (WSI). M&M: Native renal biopsy glass slides with H&E, PAS, Masson trichrome and PASM stains from 50 retrospective cases of 2025 were retrieved from tertiary care centre in north India. Slides were semi-quantitatively assessed on routine light microscopy without clinical details and Immuno-fluorescence findings. Glass slides of same cases were digitized with Hamamatsu nanozoomer digital slide scanner. After two months of wash off period WSI images of same cases with all stains were semi quantitatively assessed for all four compartments ie glomerular, tubules, interstitium and blood vessel. Results of WSI findings were compared and correlated with routine microscopy findings.

**Results:**
Out of 50 cases compartmentwise morphological diagnoses on routine LM were glomerular predominant 40, tubulo-interstitial 4 and mixed in 5 cases. No discrepancy in WSI and routine microscopy findings was observed. Almost complete concordant findings were noted in 41 cases, for global and segmental glomerulosclerosis, mesangial proliferation, crescents, assessment of IFTA, TIN, hyalinosis and sclerosis of vessels. Minor discordant findings were in 6 cases with missing mild mesangial increase in 2, wire loop lesion in 1 and exudative lesion in 1 and crescents in 2 case with WSI. Discordant morphological finding with change in morphological diagnosis in 3, with missing epimembranous spikes on WSI for membranous nephropathy (MN) in two & missing segmental sclerosis for FSGS in one case. Discussion: Discrepancy were mainly in diagnosing early MN. Intricacies of GBM were not so well appreciated on WSI, however morphological lesions were identifiable in majority cases.

**Conclusion:**
WSI provide equally good solution for majority native renal biopsy morphological diagnosis and reporting with advantage of teleconsultation, teleteaching. In case of doubt may be supplemented with routine light microscopy.

# 211 - Artificial Intelligence and Digital Pathology for Tumor-Infiltrating Lymphocyte Quantification in Breast Cancer

**Mousumi Kandangkel[1], Anirudha K Puntambekar[2], Nutan S Jumle[3], Shashank Gupta[4], Vedant Ingawale[5], Lingadahalli S Shashidhara[1,5,6], Chaitanyanand B Koppiker[7], Nameeta Shah[4], Madhura Kulkarni[7]**

[1] Indian Institute of Science Education and Research (IISER) Pune
[2] Sahyadri Hospitals Private Limited
[3] Department of Pathology, Jehangir Hospital
[4] Amaranth Medical Analyst
[5] Ashoka University, Sonipat
[6] National Center for Biological Sciences
[7] Prashanti Cancer Care Mission (PCCM)

**Background:**
Tumor-infiltrating lymphocytes (TILs) are key immune biomarkers in breast cancer, with significant prognostic value. At present, TILs are manually scored by certified pathologists, resulting in inter-observer variability that limits reproducibility and scalability. In this validation study in a distinct Indian cohort, we compare the outputs of an automated AI/ML pipeline for stromal TIL quantification against independent evaluations by two expert pathologists, using both histopathology slides and whole-slide scans.

**Methods:**
The cohort (n=355) comprised treatment-naïve FFPE breast tumors from a single centre in India. H&E-stained whole-slide images were processed with a two-stage AI pipeline for tissue segmentation and nuclear classification. Stromal TIL density (TILs/mm²) and area-based %TILs were computed. Spearman's correlation assessed inter-observer agreement for %tumor area and %TILs, concordance across breast cancer subtypes (ER+, HER2+, and TNBC), and alignment between each pathologist's %TIL scores and AI-derived stromal TIL metrics.

**Results:**
Inter-pathologist agreement was moderate for percent tumor area ($r=0.51$; $p<0.0001$) and percent TILs ($r=0.59$; $p<0.0001$). Subtype-wise percent TILs concordance remained consistent across ER+ ($r=0.49$), HER2+ ($r=0.56$), and TNBC ($r=0.50$) cases (all $p<0.0001$). AI-estimated cellular tumor area correlated positively with pathologists' percent tumor ($r=0.32–0.43$; $p<0.0001$). Automated stromal TIL metrics showed moderate to strong correlation with the pathologist's percent TILs ($r=0.52$, $p<0.0001$) ($r=0.69$, $p<0.0001$), slightly better than within the pathologist's concordance.

**Conclusion:**
The AI/ML pipeline provides automated, consistent, reproducible quantification of stromal TIL percentages with good agreement with expert pathologists. Computational TIL assessment offers a scalable, objective alternative to manual scoring and supports integration into routine digital pathology workflows.

## 212 - Learning with noisy labels for histological stain classification

**Hrafn Weishaupt[1,2], Yvan Samir Belakebi-Joly[1], Maya Maya Barbosa Silva[1,2], Stein Ivar Hallan[3,4], Rogier van Wijck[5], Sabine Leh[1,2]**

[1] Department of Pathology, Haukeland University Hospital
[2] Department of Clinical Medicine, University of Bergen
[3] Department of Renal Medicine, St. Olav University Hospital
[4] Department of Cancer Research and Molecular Medicine, Norwegian University of Science and Technology
[5] Department for Pathology and Clinical Bioinformatics, Erasmus MC University Hospital

**Background:**
Large-scale whole slide image datasets are essential for artificial intelligence (AI) development in digital pathology, but their quality is often compromised by labeling errors. In nephropathology, mislabeling of histological stains and inclusion of slides with poor quality can introduce substantial noise and bias, limiting the development of robust models. This study addresses these challenges using a learning with noisy labels framework to develop a deep learning model for histological stain classification despite imperfect training labels.

**Methods:**
The study included more than 13,000 whole slide images representing over eight histological stain categories, as well as artifacts and out-of-distribution stains. Model development followed a two-stage noisy-label learning strategy. First, the model was trained on the full dataset, including potentially mislabeled samples, to learn general stain-specific characteristics. Second, model predictions were used to identify subsets of images likely to contain clean labels, which were then used for further training. Performance was evaluated on both an internal test set and an external dataset covering most target stains.

**Results:**
Preliminary results demonstrated agreement above 97% on both internal and external test datasets. Review of the <3% discordant images indicated that most disagreements were attributable to true stain mislabeling or low-quality images. These findings confirm the presence of label noise in the test data while also demonstrating the model's ability to identify unreliable labels. Ongoing work includes detailed pathologist review of concordant and discordant predictions to further characterize model performance.

**Conclusions:**
This study demonstrates the feasibility of training a histological stain classifier using a learning with noisy labels framework. The approach enables robust model development despite mislabeled training data and supports separation of clean and noisy stain labels. The model shows potential for retrospective dataset cleaning and prospective quality assurance after slide scanning. Future work will focus on extended validation and assessment of laboratory implementation.

# 215 - Are Scanner Domain Shift Effects Histologic Pattern-Specific? Preliminary Evidence from AI-Based Lung Adenocarcinoma Subtyping

**Frederico Gaia Costa da Silva, Junya Fukuoka**
Nagasaki Univeristy

**Background:**
Artificial intelligence (AI) models in digital pathology are often developed using images from a single whole-slide scanner, raising concerns regarding scanner-induced domain shift. Whether scanner effects differentially affect specific histologic patterns remains unclear. We evaluated this question using two independent lung adenocarcinoma subtyping models.

**Methods:**
Ten lung adenocarcinoma resection specimens were digitized using five whole-slide scanners (Aperio, Hamamatsu, KFBio, Motic, and Phillips). Two independently-developed AI models trained on Aperio-scanned images were applied to all scans. AI-derived histologic subtype proportions and heatmaps were extracted for each case. Cross-scanner reproducibility was assessed using the intraclass correlation coefficient(ICC2) for subtype proportions, Fleiss' kappa for overall agreement, and pairwise Cohen's kappa for predominant subtype classification.

**Results:**
Cross-scanner reproducibility varied substantially by histologic subtype. In both AI models, mucinous (ICC2 0.838–0.895) and solid patterns (ICC2 0.817–0.867) demonstrated high reproducibility, whereas acinar (ICC2 0.327–0.451) and papillary patterns (ICC2 0.297–0.416) showed substantially lower values. Overall agreement for predominant subtype classification remained limited (Fleiss' $\kappa$=0.186 and 0.258). Pairwise analyses revealed moderate concordance among Aperio, Hamamatsu, KFBio, and Phillips scanners ($\kappa$ up to 0.72). In contrast, one scanner consistently demonstrated poor agreement with other platforms ($\kappa \approx 0$–0.25) and a marked shift toward papillary classification across both AI models, suggesting a directional scanner-specific domain shift.

**Conclusion:**
These preliminary findings suggest that scanner-induced domain shift may not simply reduce reproducibility but can alter AI-derived histologic subtype classification in a systematic, pattern-specific manner. The consistency of these effects across independent AI models supports a potential role for image acquisition characteristics in shaping AI predictions and highlights the need for multi-scanner validation in digital pathology.

# 216 - Digital Image Analysis Identifies Tissue Section Thickness as a Critical Determinant of PAX5 Immunostaining Intensity in Classic Hodgkin Lymphoma

**Kaniyappan Nambiar, Diya Ayisha, Rakhee Kar**
Jawaharlal Institute Of Postgraduate Medical Education And Research, JIPMER

**Background:**
The impact of tissue section thickness on diagnostically relevant PAX5 immunostaining in classic Hodgkin lymphoma (CHL) has not been quantitatively evaluated using spatially localized digital image analysis.

**Objective:**
To quantify the impact of tissue section thickness on PAX5 immunostaining intensity in CHL by comparing staining intensity in Reed-Sternberg cells and non-neoplastic B lymphocytes between tissue sections cut at 2 µm and 4 µm using computer-assisted image analysis.

**Methods:**
Thirty-three archived cases of classical Hodgkin lymphoma were reviewed and selected. De-identified, metadata-free regions of interest from Leica Aperio AT2-scanned images were analyzed using the online image analysis platform Trigit (trigit.com.au). RGB values were recorded from 25 randomly selected PAX5-positive Reed–Sternberg cell nuclei and 25 non-neoplastic B-cell nuclei per case in both 2-µm and 4-µm sections (825 nuclei/cell type/section thickness). Mean staining intensity (MSI) was calculated from RGB values, and luminance was derived using a standard RGB-weighted formula. MSI and luminance values between section thicknesses were compared using paired statistical analysis.

**Results:**
PAX5 staining intensity was significantly higher in 2-µm than 4-µm sections across all parameters (paired t-test, $p<0.001$). In Reed–Sternberg cells, mean staining intensity (MSI) decreased from 203.45±9.36 to 180.09±14.53, while mean luminance decreased from 201.08±9.19 to 177.57±14.24, representing reductions of approximately 23.4 and 23.5 units, respectively. In non-neoplastic B lymphocytes, MSI declined from 143.04±9.14 to 128.67±8.42, and luminance from 140.08±9.21 to 125.39±8.52, corresponding to reductions of 14.4 and 14.7 units, respectively.

**Conclusion:**
Digital image analysis identified tissue section thickness as a critical determinant of PAX5 immunostaining intensity in CHL. Compared with 4-µm sections, 2-µm sections demonstrated consistently higher and more reproducible staining in both Reed–Sternberg cells and non-neoplastic B lymphocytes. These findings provide quantitative evidence for standardizing section thickness in PAX5 immunohistochemistry and highlight the importance of pre-analytical optimization for reproducible digital pathology–based diagnostic assessment.

## 217 - Current Status of Digital PD-L1 Quantification Softwares

**Yosep Chong[1], Rajni Yadav[2], Jennifer Kim[3], Rajiv Kaushal[4], Akansha T[4], Özben Yalçın[5], Sülen Sulen[6]**
[1] College of Medicine, The Catholic University of Korea, South Korea
[2] AIIMS
[3] University of Sydney
[4] Tata Memorial Hospital
[5] Sağlık Bilimleri Üniversitesi, Prof. Dr. Cemil Taşçıoğlu Şehir Hastanesi
[6] Izmir University of Economics

Background. Programmed death-ligand 1 (PD-L1) immunohistochemistry (IHC) is the most widely used companion diagnostic biomarker in immuno-oncology, guiding decisions for more than twenty FDA-approved indications across ten cancer types. Yet PD-L1 scoring suffers from reproducibility problems that directly affect patient access to immune checkpoint inhibitor (ICI) therapy: inter-observer agreement is substantial for some cutoffs but drops sharply near the clinically critical 1% and 50% thresholds, and results depend on the specific antibody clone, scoring system (TPS, CPS, IC), and tumor type. Artificial intelligence (AI)–assisted morphometry has emerged as a promising solution, but its regulatory status, clinical performance, and optimal deployment model remain unclear. Scope. We synthesize evidence from 54 peer-reviewed studies covering rule-based image analysis, supervised deep learning, weakly-supervised and Vision Transformer architectures, H&E–based prediction, and reader-with-AI workflows. Key findings. (i) Modern AI algorithms achieve intraclass correlation coefficients of 0.90–0.98 in controlled NSCLC cohorts, with PathAI reporting AI performance (ICC 0.984) exceeding a 24-pathologist panel (ICC 0.952). (ii) Real-world validation using multi-clone cross-comparison shows that AI can correct systematic pathologist under-scoring rather than merely introduce errors. (iii) Reader-with-AI workflows show the largest reproducibility gains on hardest cases (Haragan 2024 Fleiss κ 0.613→0.886; Ito 2024 difficult-subset κ 0.52→0.78). (iv) H&E-based PD-L1 prediction is maturing rapidly (AUC 0.88–0.93). (v) No PD-L1 AI has received FDA clearance as of 2026, though five CE-IVD products are in clinical use in Europe. Conclusions. Evidence supports reader-with-AI deployment as the near-term clinical pathway, with full AI replacement remaining a long-term research frontier contingent on prospective randomized trials, regulatory precedent, and liability frameworks. Remaining priorities include cross-clone harmonization, performance at the 1% cutoff, and validation in underserved cancer types such as cervical, ovarian, and the 2026-approved all-histology esophageal indication.

# 220 - Predicting Mismatch Repair Status in Colorectal Adenocarcinoma from Whole-Slide Images: Development and Validation of a Pathologist-Guided Deep Learning Model

**Bharathi Prabakaran Ns[1], Rajni Yadav[2], Priyanshu Sharma[3], Amit Yadav[3]**
[1] All India Institute Of Medical Sciences
[2] All India Institute of Medical Sciences
[3] Translational Health Science and Technology Institute

**Background/Objectives:**
Mismatch repair (MMR) status is an important biomarker in colorectal adenocarcinoma with significant therapeutic and prognostic implications. Assessment is routinely performed using immunohistochemistry (IHC), which may not always be readily available in resource-limited settings. This study aimed to develop and evaluate a deep learning model capable of predicting MMR status directly from digitized Hematoxylin and eosin (H&E)-stained whole-slide images (WSIs).

**Methods:**
An initial pilot phase was conducted using approximately 80 IHC-characterized colorectal adenocarcinoma WSIs to explore the feasibility of image-based MMR prediction. Based on these findings, a larger cohort of 650 WSIs was assembled for model development. Tumour-rich regions, including the tumour microenvironment, were manually annotated by pathologists, while artefacts and non-informative areas were excluded. Approximately 400,000 image patches were generated, from which 50,000 high-quality patches were selected for training and validation. A deep learning model based on the EfficientNet-B3 architecture was trained using five-fold cross-validation. Independent testing was subsequently performed on a blinded external cohort of 150 cases comprising 75 pMMR and 75 dMMR tumours.

**Results:**
Early exploratory models achieved accuracies ranging from 58% to 72%, highlighting the value of patch-based analysis. The final model achieved an internal validation accuracy of approximately 86%. Inclusion of tumour microenvironmental features appeared to enhance predictive performance. On external testing, the model achieved an accuracy of 53.7%, sensitivity of 58.1%, specificity of 49.3%, and an area under the receiver operating characteristic curve of 0.54. Reduced performance in the external cohort suggested the influence of tumour heterogeneity and inter-institutional variation in slide preparation and image acquisition.

**Conclusion:**
This study demonstrates the feasibility of predicting MMR status from routine H&E-stained WSIs using a pathologist-guided deep learning approach. Although internal performance was encouraging, the decline in performance during external validation highlights the need for larger multicentre datasets and further refinement before clinical implementation.

## 221 - CorePath: a real-world-validated agentic multimodal whole-slide foundation-model system for colorectal pathology

**Yihui Wang, Hao Chen**
The Hong Kong University Of Science And Technology

Colorectal cancer (CRC) management relies heavily on histopathological assessment of whole-slide images (WSIs) across a broad clinical workflow, spanning endoscopic biopsy assessment, resection reporting, molecular biomarker evaluation and prognostic decision-making. However, routine pathology faces major bottlenecks, including high workload, inter-observer variability, and reliance on costly ancillary testing for molecular biomarkers. Existing AI systems often struggle with narrow task scopes and lack large-scale prospective validation. Here we introduce CorePath, an integrated colorectal pathology system comprising two core components: CorePath-FM, a visual foundation model, and CorePath-MLLM, a WSI-level multimodal large language model. We assembled a diverse repository of 156,325 slides from 14 private centers and several public datasets, curating a subset of 65,177 WSIs (280 million patches). This large-scale cohort supported the visual foundation training of CorePath-FM and provided 300K instruction-tuning samples for CorePath-MLLM to generate structured, evidence-linked diagnostic summaries. We further connected these components with task-specific predictors through an agentic decision-support framework for case-level interpretation and workflow-aware recommendations for pathologist review. We benchmarked CorePath on 30 workflow-aligned tasks across 155 datasets, including a five-center prospective cohort. CorePath-FM ranked first in 152 of 155 evaluations (mean rank: 1.02), indicating consistent superiority. In prospective validation, it ranked first across all 20 dataset-level evaluations with a mean AUC of 0.8949, proving robust real-world generalization. In deployment analyses, CorePath's agentic workflow facilitated sensitivity-prioritized clinical triage. It successfully auto-ruled out 65% of biopsy cases from pathologist review with zero missed malignancies and reduced MSI/dMMR reflex testing by 60%. Finally, a multi-reader crossover study involving eight pathologists across three tasks demonstrated that AI assistance significantly improved diagnostic accuracy, reduced reading time by 19.01%, increased diagnostic confidence by 11.85%, and improved inter-reader agreement. Ultimately, CorePath successfully combines visual foundation modeling, multimodal interpretation, and agentic orchestration to deliver highly effective, evidence-aware human-AI decision support.

# 222 - Generative AI-Based Virtual IHC Staining and Subtype Classification in Breast Cancer: A Comparative Study of H&E-to-IHC Translation versus Direct H&E Prediction

**Svetlana Illarionova[1,2], Anastasiia Studenikina[1,2], Aida Akaeva[1,2], Fedor Gubanov[1,2], Rifat Hamoudi[2], Maxim Sharaev[1,2]**

[1] Skolkovo Institute Of Science And Technology

[2] University of Sharjah

**Background/Objective:**

Immunohistochemistry (IHC) for estrogen receptor (ER) and human epidermal growth factor receptor 2 (HER2) is essential for breast cancer subtyping and treatment selection. However, physical IHC is costly, slow, and tissue-consuming. Generative AI enables virtual staining from routine H&E sections. This study compares two machine learning strategies: (1) direct classification of ER/HER2 expression scores from H&E, and (2) a two-step AI pipeline (virtual staining from H&E to IHC and further expression scoring). We evaluate virtual IHC fidelity and classification accuracy against ground-truth physical IHC.

**Methods:**

A diffusion- and GAN-based models were trained on 100 breast cancer patients with paired H&E and IHC whole-slide images for ER and HER2. Virtual IHC was generated from H&E. Expression scores (positive/negative and semiquantitative for HER2 0–3+) were predicted using: (i) direct classification from H&E (Virchow2 foundation model), and (ii) a separate classifier from virtual IHC. Ground truth was established by pathologists based on physical IHC. Generative quality was assessed with the histology-aware metrics; classification performance was measured by AUC and F1.

**Results:**

Direct H&E classification (Virchow2) achieved AUC/F1 of 86.9/79.1 for ER and 89.1/64.3 for HER2. The two-step generative approach yielded comparable or higher HER2 classification results and improved ER classification. Virtual IHC achieved high perceptual alignment with expert scores. For HER2 scoring, the two-step method reduced ambiguity in intermediate cases and provided interpretable membrane patterns. ER nuclear expression was accurately preserved.

**Conclusion:**

Generative AI-based virtual IHC staining for ER and HER2 offers a clinically viable alternative to physical IHC, matching or surpassing direct H&E classification for HER2 while adding pathologist-interpretable visual output. This approach is particularly promising for rapid, resource-saving molecular diagnostics in digital pathology workflows.

# 225 - Development of an AI-Based Model for Classifying Malignant and Benign Breast Tumors from Digital Histopathology Images

**Tradygta Chantya Raissa Natanggi[1], Alif Muhammad Hafizh[2], Tantri Hellyanti[3], Dyah Ayu Woro Setyaningrum[4], Ilva Widyaningtyas[3], Prasandhya Astagiri Yusuf[2,5]**

[1] Faculty of Medicine Universitas Indonesia
[2] Health Big Data Center IMERI, Faculty of Medicine Universitas Indonesia
[3] Department of Pathology Anatomy Dr. Cipto Mangunkusumo National Central Public Hospital / Faculty of Medicine Universitas Indonesia
[4] Department of Pathology Anatomy, Faculty of Medicine Universitas Trisakti
[5] Department of Medical Physiology and Biophysics, Faculty of Medicine Universitas Indonesia

**Background/Objective**
Breast cancer remains a major cause of cancer-related mortality among women in Indonesia. Histopathological diagnosis still relies heavily on manual interpretation by pathologists, which may affect efficiency and diagnostic consistency. This study aimed to develop a deep learning model for classifying breast tumors from whole-slide histopathology images using locally sourced Indonesian data.

**Methods**
A quantitative experimental study was conducted using a Vision Transformer-based deep learning model trained on hematoxylin-eosin (H&E) whole-slide images of breast tumors. The dataset comprised 127 digital histopathology slides that underwent quality control and expert annotation by anatomical pathologists to establish ground-truth labels. Following annotation, image data were processed through augmentation, stain normalization, and domain adaptation to reduce inter-institutional variability. Model optimization included patch embedding and token representation tuning. Performance was evaluated using accuracy, sensitivity, specificity, F1-score, and area under the receiver operating characteristic curve (AUC). Model interpretability was assessed using Gradient-weighted Class Activation Mapping (Grad-CAM).

**Results**
Preliminary analysis was performed using histopathologically confirmed invasive ductal carcinoma cases. The Vision Transformer model achieved a training accuracy of 0.858 and a validation accuracy of 0.921. Grad-CAM visualization showed consistent attention to diagnostically relevant tissue regions. Stain normalization and domain adaptation also reduced visual variability across slides obtained from different institutions. Collection and analysis of benign breast tumor images are ongoing to complete the final classification model.

**Conclusion**
These preliminary findings suggest that Vision Transformer-based models can accurately identify invasive ductal carcinoma patterns from digital histopathology images. The combination of explainable AI, stain normalization, and domain adaptation supports model interpretability and robustness across datasets. This approach demonstrates the potential of AI-assisted digital pathology to support breast tumor classification in Indonesia, while ongoing integration of benign tumor data is expected to further refine model performance. Keywords: artificial intelligence, deep learning, breast cancer, whole slide imaging, explainable AI

# 226 - Efficiency of Digital Intradepartmental Consultation in a Multi-Center Pathology Network: Analysis of 5,882 Consultations

**Fatma Yildirim, Sulen Sarioglu, Burcu Cekic, İlknur Turkmen, Serdar Balci**
Memorial Healthcare Group, Department of Pathology

**Introduction:**
Intradepartmental consultation (IC) is a key quality-assurance process in surgical pathology, supporting diagnostic accuracy through expert review. Fully digital pathology systems provide comprehensive audit trails that enable objective evaluation of consultation workflows. We assessed the efficiency, utilization patterns, workload distribution, and network characteristics of IC within a large multi-center digital pathology network. Materials and Methods: A retrospective study was conducted across a fully digital whole-slide imaging network comprising 15 hospital sites. Intradepartmental consultation records from August 2022 to December 2025 were analyzed. Consultation frequency, turnaround time (TAT), temporal patterns, workload distribution, and network structure were evaluated using a reproducible R-based workflow.

**Results:**
A total of 5,882 consultations involving 4,532 unique cases and 33 pathologists were analyzed over a 39.4-month period. IC was requested in approximately 1.9% of all pathology cases. Median TAT was 3.0 hours (IQR 1.0–14.8), while mean TAT was 10.3 hours (SD 16.1). Overall, 88.9%, 95.5%, and 98.5% of consultations were completed within 24, 48, and 72 hours, respectively. Consultations initiated during weekdays demonstrated substantially shorter TATs than those initiated during weekends (2.7 vs. 11.8 hours), and morning consultations were completed faster than evening consultations (2.1 vs. 14.2 hours). The most frequent consultation requests involved inflammatory/non-neoplastic lesions (26.8%), diagnosis/tumor classification (20.5%), and dysplasia/grade assessment (16.0%). Network analysis identified 33 nodes connected by 459 directed consultation links organized into three subspecialty-oriented communities. Workload distribution was markedly unequal (Gini coefficient = 0.621), indicating concentration of consultation activity among a limited number of expert pathologists.

**Conclusion:**
Digital pathology enables highly efficient intradepartmental consultation across geographically distributed centers, achieving rapid turnaround times and supporting structured expertise sharing. Continuous monitoring of consultation metrics may facilitate expertise-based case routing, workload balancing, and further optimization of diagnostic workflows.

# 227 - Classifying Malignant and Benign Breast Tumors from Digital Histopathology Images Using Artificial Intelligence: A Scoping Review

**Tradygta Chantya Raissa Natanggi[1], Listya Tresnanti Mirtha[2,3]**

[1] Faculty of Medicine Universitas Indonesia

[2] Sports Medicine Residency Training Program, Faculty of Medicine Universitas Indonesia

[3] Medical Technology Cluster, Indonesian Medical Education and Research Institute, Faculty of Medicine Universitas Indonesia

**Background/Objective**

Breast cancer remains one of the leading causes of morbidity and mortality among women worldwide, making accurate and efficient tumor classification increasingly important. Digital histopathology has become a central modality for breast cancer diagnosis because it directly visualizes tissue morphology and cellular architecture. However, manual interpretation by pathologists may be time-consuming and subject to interobserver variability. Therefore, this scoping review aimed to analyze the development of artificial intelligence (AI)-based methods for classifying malignant and benign breast tumors using digital histopathology images, identify the most effective AI approaches, and evaluate existing research gaps and future opportunities in this field.

**Methods**

This study used a scoping review design based on ten international journal articles related to AI, machine learning, and deep learning applications in breast cancer classification. The reviewed studies included histopathology-based classification models, mammography, MRI, ultrasound elastography, and umbrella review evidence. Data were collected using a structured extraction matrix containing study characteristics, AI methods, performance outcomes, limitations, and research relevance. Narrative thematic synthesis was used to analyze and compare findings across studies.

**Results**

The findings showed that deep learning-based approaches, particularly convolutional neural networks, transfer learning, ensemble learning, and hybrid CNN-LSTM models, demonstrated the highest classification performance. Several studies reported classification accuracy above 98% for benign and malignant breast tumor classification using digital histopathology images. The studies also identified important limitations, including dataset dependency, preprocessing variability, limited explainability, computational burden, and insufficient clinical validation.

**Conclusion**

Artificial intelligence shows strong potential for improving breast tumor classification using digital histopathology images. Deep learning approaches currently represent the most effective methods for benign and malignant classification. Future studies should focus on improving generalizability, explainability, multicenter validation, and clinical applicability to support the advancement of digital pathology and precision medicine. Keywords: artificial intelligence, deep learning, breast cancer, classification, digital pathology

# 228 - Immunohistochemistry Utilization Patterns Across Pathology Subspecialties: A Data Mining-Based Analysis of a Comprehensive Pathology Information System

**Fatma Yildirim, Sulen Sarioglu, İlknur Turkmen, Serdar Balci**
Memorial Healthcare Group, Department of Pathology

**Background/Objective:**
This study aims to characterize immunohistochemical (IHC) antibody utilization patterns across pathology subspecialties through systematic data mining of an Anatomic Pathology Laboratory Information System (APLIS), with the goal of identifying subspecialty-specific IHC profiles, panel co-application patterns, temporal utilization trends, and parameters for evidence-based resource planning. Material and Methods: IHC utilization data were retrospectively extracted and standardized via data mining from a APLIS encompassing 20 subspecialty sections over a five-year period. Of 352,308 total cases, 43,629 (12.4%) had IHC applied and were analyzed for per-case antibody counts, subspecialty-specific frequencies, co-application pairs, and year-over-year trends.

**Results:**
The subspecialties with the highest IHC case volumes were Gynecologic Pathology (n=6,586; 15.1%), Gastrointestinal Pathology (n=5,187; 11.9%), and Breast Pathology (n=5,016; 11.5%). The leading antibodies in Gynecologic Pathology were p16 (76.7%), Ki-67 (19.6%), and p53 (12.6%). In Gastrointestinal Pathology, the most frequently applied markers were CMV (19.7%), p53 (18.0%), and CK7 (16.6%), alongside neuroendocrine markers and MMR proteins. Breast Pathology demonstrated near-universal application of the predictive panel: ER (86.1%), PR (73.9%), Ki-67 (71.4%), and HER2 (70.1%). Hematopathology recorded the highest per-case antibody intensity (mean 8.9 antibodies/case). The most frequently co-applied antibody pairs were ER+PR (n=5,291), CD20+CD3 (n=5,174), and ER+Ki-67 (n=4,718), reflecting standardized diagnostic and predictive panel practice. Temporal trend analysis revealed a consistent post-2022 increase across all major subspecialties: CMV and neuroendocrine markers showed the steepest rise in Gastrointestinal Pathology, myeloid markers (CD61, MPO) emerged prominently in Hematopathology, and all four predictive breast markers demonstrated sustained parallel growth.

**Conclusion:**
Systematic data mining of APLIS enables high-resolution mapping of IHC utilization patterns, revealing distinct subspecialty-specific antibody profiles, standardized panel behaviors, and clinically meaningful temporal trends. The observed post-2022 surge underscores the expanding role of predictive IHC, and provides a robust framework for evidence-based laboratory planning and resource optimization.

# 230 - Automated Tissue Core Identification and Segmentation in Multi-Core Histology Whole-Slide Images

**Sean Lim[1], Chee Leong Cheng[2], Li Yan Khor[2], Ye Heng Seow[2]**
[1] Lee Kong Chian School of Medicine
[2] Anatomical Pathology, Singapore General Hospital

**Background/Objective:**
It is becoming more common place for laboratories to place multiple biopsy cores on a single glass slide to reduce costs of preparation, staining, and scanning. However, this creates a major barrier for downstream AI, which typically assumes a single core per slide. It is important to separate and label individual tissue cores for the downstream AI models to detect and grade any cancers for localised therapy. We aimed to develop and evaluate an AI system that automatically identifies, groups, and labels individual biopsy cores on multi-core whole-slide images (WSIs) to support subsequent digital pathology workflows.

**Methods:**
Ground truth core masks were created by combining all fragments belonging to the same ink colour into a single region of interest. These were checked against the outputted mask from our algorithm. Our pipeline comprises of 4 steps. (1) WSI downsampling and tissue masking; (2) generation and simplification of bounding boxes around tissue fragments; (3) clustering of fragments by spatial proximity and colour to reconstruct whole cores; and (4) export of per-core ROIs in a machine-readable format for integration with downstream AI models.

**Results:**
The system successfully processed multi-core slides including highly fragmented and closely apposed cores. Core grouping by colour and spatial clustering produced consistent per-core masks suitable for subsequent cancer detection and grading. The output aligned with laboratory labelling conventions (e.g. target vs saturation series). Additionally, the system runs fully on algorithms, allowing it to run offline and be more easily interpretable and improved on than complex AI approaches like neural networks

**Conclusion:**
We present an automated model that identifies and segments individual biopsy cores on multi-core histology. This core-level preprocessing layer addresses a key bottleneck in deploying AI to routine multi-core biopsy workflows and enables more accurate, scalable, and cost-efficient digital pathology analysis.

# 231 - INTERPATH: A Workflow-Integrated Digital Pathology Platform for Remote Consultation

**Youngjin Kang, Sangjeong Ahn**
Korea University Anam Hospital

**Objectives**
Remote consultation is an essential component of pathology practice, particularly for diagnostically challenging cases requiring subspecialty expertise. Conventional glass slide-based consultation is often associated with logistical delays. We developed INTERPATH, an in-house digital pathology platform, as a workflow-integrated alternative. This study evaluated the feasibility, turnaround time (TAT), and user satisfaction of INTERPATH in real-world consultation practice.

**Methods**
INTERPATH interfaces directly with the institutional pathology image management system, enabling seamless transfer of whole-slide images and gross photographs with automatic de-identification of patient information. Consultation requests were distributed through an integrated email function. Consultants received encrypted case links, reviewed images through a web-based viewer, and submitted opinions using an integrated commenting system. User satisfaction was assessed using the System Usability Scale (SUS).

**Results**
Twelve consultation cases were reviewed by thirteen external board-certified pathologists with relevant subspecialty expertise, including one case independently reviewed by two pathologists. Cases represented a broad spectrum of benign and malignant lesions across thoracic, genitourinary, gastrointestinal, skin, lymphoreticular, breast, gynecologic, and neuropathology. All consultants successfully accessed and reviewed their assigned cases, and consultation comments were submitted for 12 of 13 consultations. No technical failures were encountered. The mean TAT from consultation request to opinion submission was 682.8 minutes, enabling same-day consultation in most cases. All consultants completed the survey, and the mean SUS score was $90.6 \pm 8.0$, indicating exceptional usability. Survey feedback indicated that even first-time users were able to rapidly review all whole-slide images through the web-based platform.

**Conclusions**
INTERPATH successfully integrated image management, automated de-identification, consultation tracking, and communication within a unified digital workflow. The platform demonstrated excellent usability and supports scalable, interoperable deployment across diverse pathology practice settings.

# 233 - Benchmarking Vision-Language Models for Automated Pathology Diagnosis and Report Generation

**Sangjeong Ahn[1], Yumi Lee[2], Jang-Hwan Choi[2], Sulen Sarioglu[3], Serdar Balcı[3], Ilknur Turkmen[3], Junya Fukuoka[4], Andrey Bychkov[4], Jijgee Munkhdelge[4], Rajiv Kumar Kaushal[5], Rajni Yadav[6], Yuri Tolkach[7]**

[1] Korea University Anam Hospital
[2] Department of Artificial Intelligence, Ewha Womans University
[3] Department of Pathology, Memorial Health Group
[4] Department of Pathology, Kameda Medical Center
[5] Department of Pathology, Tata Memorial Hospital
[6] Department of Pathology, All India Institute Of Medical Sciences Delhi
[7] Institute of Pathology, University Hospital Cologne

The rapid advancement of vision-language models (VLMs) has accelerated progress in computational pathology; however, whole-slide image (WSI)-based pathology report generation remains limited by the scarcity of large-scale WSI–report datasets and the complexity of mapping spatially distributed visual patterns to structured clinical text. To address this, we introduce a clinically curated Pan-Asia WSI–report dataset of approximately 10,500 pairs from five institutions and establish the REG 2025 benchmark through a MICCAI challenge for systematic evaluation of multimodal models. We analyze submitted methods spanning pretrained VLMs, multiple-instance learning frameworks, hierarchical expert models, retrieval-augmented generation, and cross-modal Transformers. Rather than indicating that VLM use alone was sufficient for superior performance, the results suggest that top-performing methods benefited from structured report representations, hierarchical diagnostic decomposition, and effective multimodal grounding. We identify key limitations, including instability in quantitative attribute estimation (e.g., numeric hallucination) and a tendency toward diagnostic overspecification, with some errors resembling known diagnostic pitfalls in routine pathology. These findings establish REG 2025 as a benchmark for evaluating WSI-based structured report generation and vision-language understanding in computational pathology, providing insights for the design of clinically grounded multimodal pathology models.

# 234 - MORPHOMETRIC ANALYSIS OF SUPERFICIAL TUMOR INFILTRATIVE NESTS AND ITS CORRELATION WITH WORST PATTERN OF INVASION IN ORAL SQUAMOUS CELL CARCINOMA

**Siva Barathy R, Kalaivani selvi, Lavanya murugesu, Sruthi sakthivadivel**
SRI VENKATESHWARAA MEDICAL COLLEGE HOSPITAL AND RESEARCH CENTRE,DEPARTMENT OF PATHOLOGY

MORPHOMETRIC ANALYSIS OF SUPERFICIAL TUMOR INFILTRATIVE NESTS AND ITS CORRELATION WITH WORST PATTERN OF INVASION IN ORAL SQUAMOUS CELL CARCINOMA

## INTRODUCTION

Oral squamous cell carcinoma (OSCC) exhibits variable biological behavior despite similar histopathological grades. Tumor infiltrative nests(TINs) may indicate tumor aggressiveness. Worst pattern of invasion(WPOI)is an established prognostic factor with WPOI 4,5 associated with poorer outcomes. However, WPOI assessment requires resection specimens and cannot be reliably evaluated in superficial diagnostic biopsies. This study evaluated superficial infiltrative nest morphometry and its association with WPOI

## OBJECTIVES

To assess the area and perimeter of superficial infiltrative tumor nests in OSCC and correlate these parameters with WPOI.

## METHODS

A retrospective study was conducted on 20 histologically confirmed OSCC cases. Digital morphometric analysis of superficial infiltrative tumor nests within 2mm from the tumor surface was performed on H&E stained sections using Mosaic version 2.4. A hotspot withthe smallest nesting pattern was selected and ten consecutive nests were measured. Mean area and perimeter were correlated with WPOI.Cases with WPOI 1-3 were categorized as Group I and WPOI 4-5 as Group II.Statistical analysis was performed using Mann-Whitney U test, with $p<0.05$ considered significant.

## RESULTS

The overall mean tumor nest area was 3903.7±4084.7 and the mean perimeter was 362.95±444.37. WPOI 1-3cases(n=9) showed higher mean tumor nestarea(6219.22±5061.75) and perimeter(589.67±603.26) compared to WPOI 4-5cases(n=11),which demonstrated a mean area of 2010±1561.43 and perimeter of 177.45±57.67. Significant differences were observed forboth tumor nest area($p=0.006$)and perimeter ($p=0.016$) between the groups.

## CONCLUSION

Smaller superficial tumor nests were associated with WPOI 4-5, suggesting a more aggressive invasive pattern. Morphometric assessment of superficial tumor nests may serve as an adjunctive prognostic tool in OSCC and may aid risk stratification in diagnostic biopsies. Keywords: Oral squamous cell carcinoma; Tumor infiltrative nests; Morphometric analysis; Worst pattern of invasion; Prognosis. AUTHORS: DR.SIVA BARATHY R(PG) DR.KALAIVANI(PROFESSOR) DR.LAVANYA M(PROFESSOR&HOD) DR.SRUTHI(ASSISTANT PROFESSOR)

# 236 - Restoration of Large-Scale Tissue Artifacts in Whole Slide Images Using Context-Aware Diffusion Models

**Sangjeong Ahn[1,2], Hyeseong Lee[2], Sung Hak Lee[3]**

[1] Department of Pathology, Korea University Anam Hospital, College of Medicine, Korea University

[2] Department of Biomedical Informatics, Korea University College of Medicine

[3] Department of Hospital Pathology, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Kore

**Background / Objective:**
While deep learning has significantly advanced Whole Slide Image (WSI) analysis, common tissue artifacts such as bubbles and folds often induce silent failures by obscuring critical cellular morphology. Existing image restoration methods are predominantly constrained to small patch levels, failing to preserve global structural coherence across large-scale slides. To address this, we present a novel framework designed for globally consistent, large-scale artifact inpainting to recover diagnostic structures and prevent high-confidence misclassifications.

**Methods:**
The proposed method leverages a Latent Diffusion Model constrained by Pathology Foundation Model embeddings. To effectively capture long-range dependencies during random patch diffusion across extensive artifact regions, Large Kernel Attention is integrated into the architecture. Furthermore, Distance-Weighted Interpolation and an Adaptive Guidance Scale are utilized to seamlessly modulate contextual information from surrounding regions, ensuring high-fidelity structural consistency across megapixel-scale transitions.

**Results:**
Extensive evaluations on TCGA-BRCA, BACH, and Camelyon16 datasets, demonstrating robust scalability on high-resolution tissue environments exceeding dimensions of 2048X2048 pixels up to 4608 pixels. The proposed model significantly improved image quality, reducing the Fréchet Inception Distance to 29.66 on TCGA-BRCA. In downstream analysis on the BACH dataset, the Adjusted Rand Index (ARI) score improved from 0.1256 to 0.3656 after restoration, recovering toward the origin score of 0.5822. Furthermore, our framework achieved a ResNet50 classification accuracy of 75.5%±1.3% on BACH, outperforming the raw artifact baseline 61.0%±3.7% and the conventional detect-and-discard pipeline 64.2%±4.1%. For WSI-level evaluation on Camelyon16, the model restored ABMIL classification accuracy to 84.0%±1.2% and achieved an AUC of 90.3%±1.9%, demonstrating a robust improvement over traditional quality control procedures.

**Conclusion:**
By ensuring large-scale structural integrity without compromising macro-level tissue context, this diffusion-based approach effectively mitigates performance degradation caused by preparation artifacts. The framework preserves critical diagnostic morphology, thereby offering a robust pre-processing alternative to improve the reliability of clinical AI models in digital pathology.