P45

Towards Tumor-Agnostic IHC Cell Detection and Classification

Elin Samuelsson[1], Jean-Eudes Le Douget[1], Alexandre Filiot[1], Jérémy Guillon[1], Benoit Schmauch[1], David Lin[1], Cécile Maisin[2], Kim Vianey[2], Mina Farag[1], Corentin Wiscart[1], Hortense Deslandes[1], Zoé Vaquette[1], Katharina von Loga[1]

[1] Waiv, formerly Owkin Dx, Paris, France. [2] Cypath & Cypath-RB, Villeurbanne, France.

Background

Cell detection and classification in immunohistochemistry (IHC) whole slide images (WSIs) remain fundamental challenges for accurate IHC scoring in digital pathology. As new IHC markers and indications expand with the advances of antibody-drug conjugates (ADCs), AI-powered IHC cell detection and classification must be robust to different tumor types, or require a limited amount of data for fine-tuning.

Objectives

Our goal was to develop a cell detection and classification model specifically tailored for IHC WSIs, by exploiting an internal IHC-specific foundation model. Performance was assessed on external cohorts, both in-domain (same indication and marker) and out-of-domain. For the latter case, we also explored few-shot fine-tuning (~20-30 tiles from three indications) for domain adaptation.

Data

Three independent cohorts were available: two breast cancer HER2 cohorts and one multi-indication (gastric, lung, prostate) multi-marker cohort. Point annotations were made by expert pathologists, and expanded to nuclei with an internal model based on Nuclick [1].

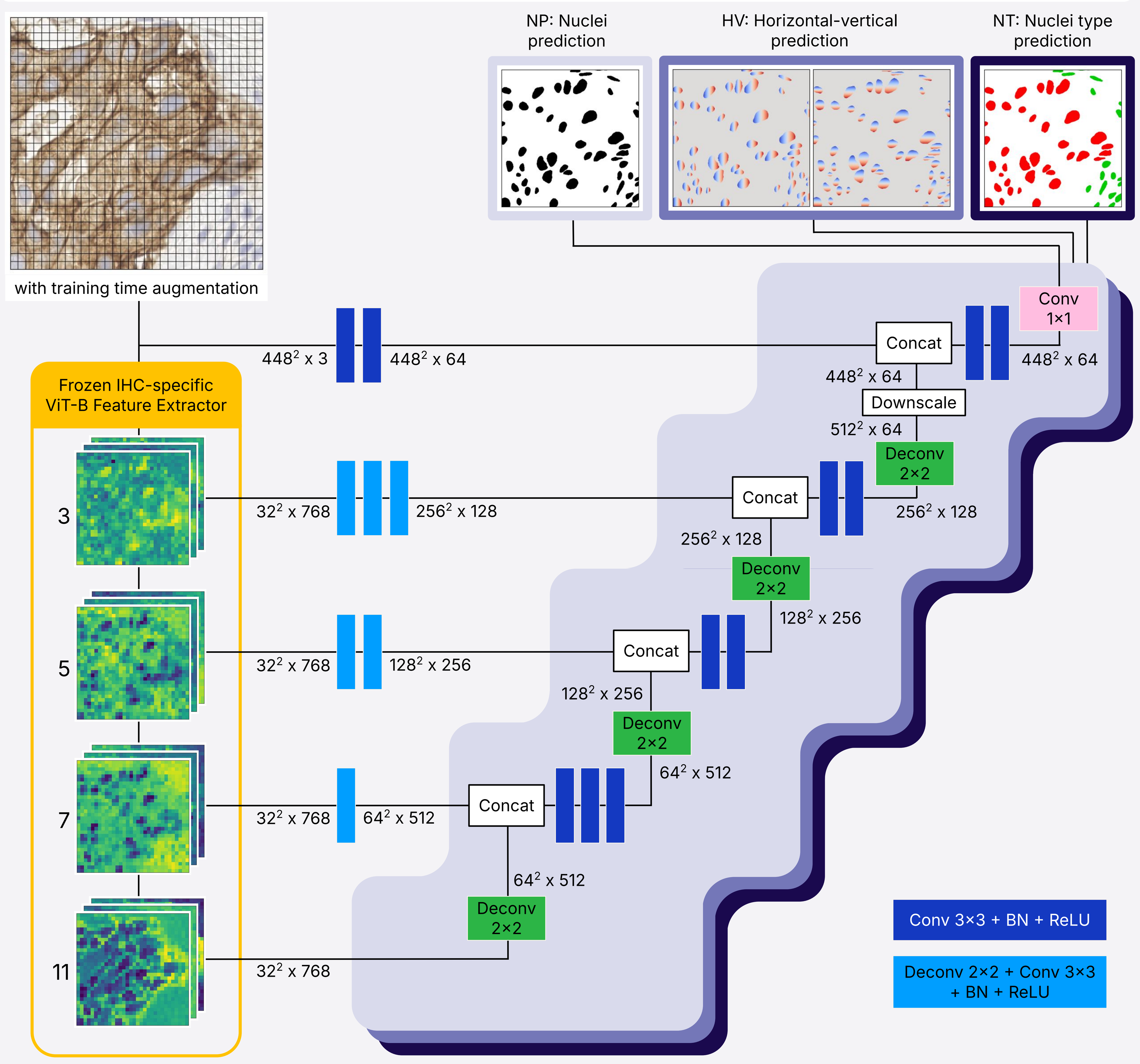
Source	Indication	Marker	WSIs	Tiles	Cells	Cancer cells	Other cells
Cypath	Breast	HER2	49	699	56,606	35,273 (62%)	21,333 (38%)
NHS	Breast	HER2	52	245	19,957	11,774 (59%)	8,183 (41%)
FTB	Gastric	CK20 / CK7	13	46	3,546	2,109 (59%)	1,437 (41%)
	Lung	PDL1	6	27	2,068	658 (32%)	1,410 (68%)
	Prostate	NKX3 / PSA	15	49	6,454	4,728 (73%)	1,726 (27%)

Methods

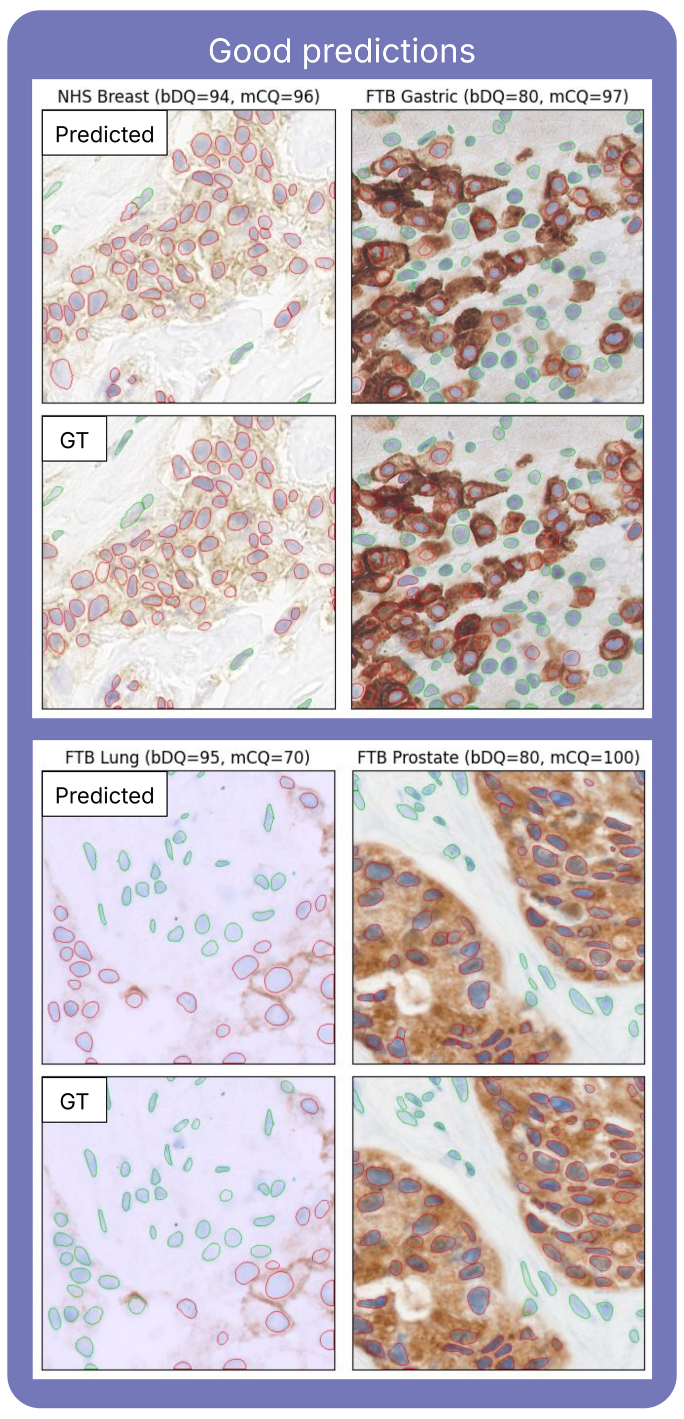
Our cell detection and classification model is based on the CellViT architecture [2] with an internal IHC-specific foundation model as frozen backbone. Leveraging the nuclei prediction (NP) map and gradients of the horizontal-vertical (HV) maps, a post-processing watershed algorithm separates cell instances. Primary training was run on the Cypath breast cancer HER2 cohort, followed by a joint few-shot fine-tuning on all FTB cohorts, using three-fold cross-validation with cohort-stratified splits.

Metrics computation relies on matching ground truth (GT) and predicted nuclei, defining pairs with intersection over union (IoU) >0.5 as true positives (TP). Remaining GT and predicted nuclei are false negatives (FN) and false positives (FP), respectively. Rather than evaluating per tile, we pool all cells from all tiles to ensure equal contribution of every cell. For each class c, TP is further divided into four subgroups based on the GT and predicted class labels: TPc, FNc, FPc, TNc (true negatives).

Panoptic quality (PQ) is defined as the product of **detection quality (DQ)**: F1 score of TP, FN, and FP counts, and **segmentation quality (SQ)**: average IoU of all TP pairs. These can be computed per class by replacing TP → TPc, FN → (FN + FNc), and FP → (FP + FPc), to also measure classification performance. We also computed a dedicated **classification quality (CQ)**: F1 score of TPc, FNc, and FPc counts. Lastly, multiclass metrics are defined as the average of the corresponding class-wise metrics.

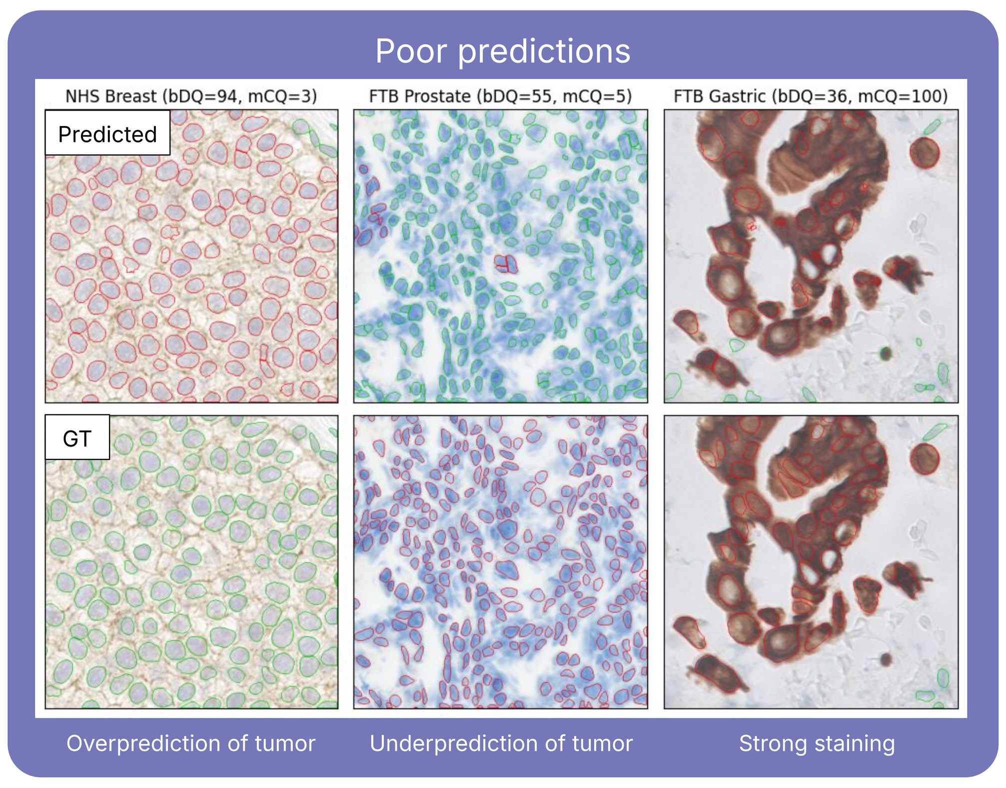


Results: Primary training



Source	Indication	Binary	Multiclass			
		DQ	CQ	DQ	SQ	PQ
NHS	Breast	84	87	73	82	60
FTB	Gastric	75	68	51	80	41
	Lung	75	61	46	78	35
	Prostate	73	70	51	77	39

External evaluation metrics of the primary model trained on the Cypath breast cancer HER2 cohort.

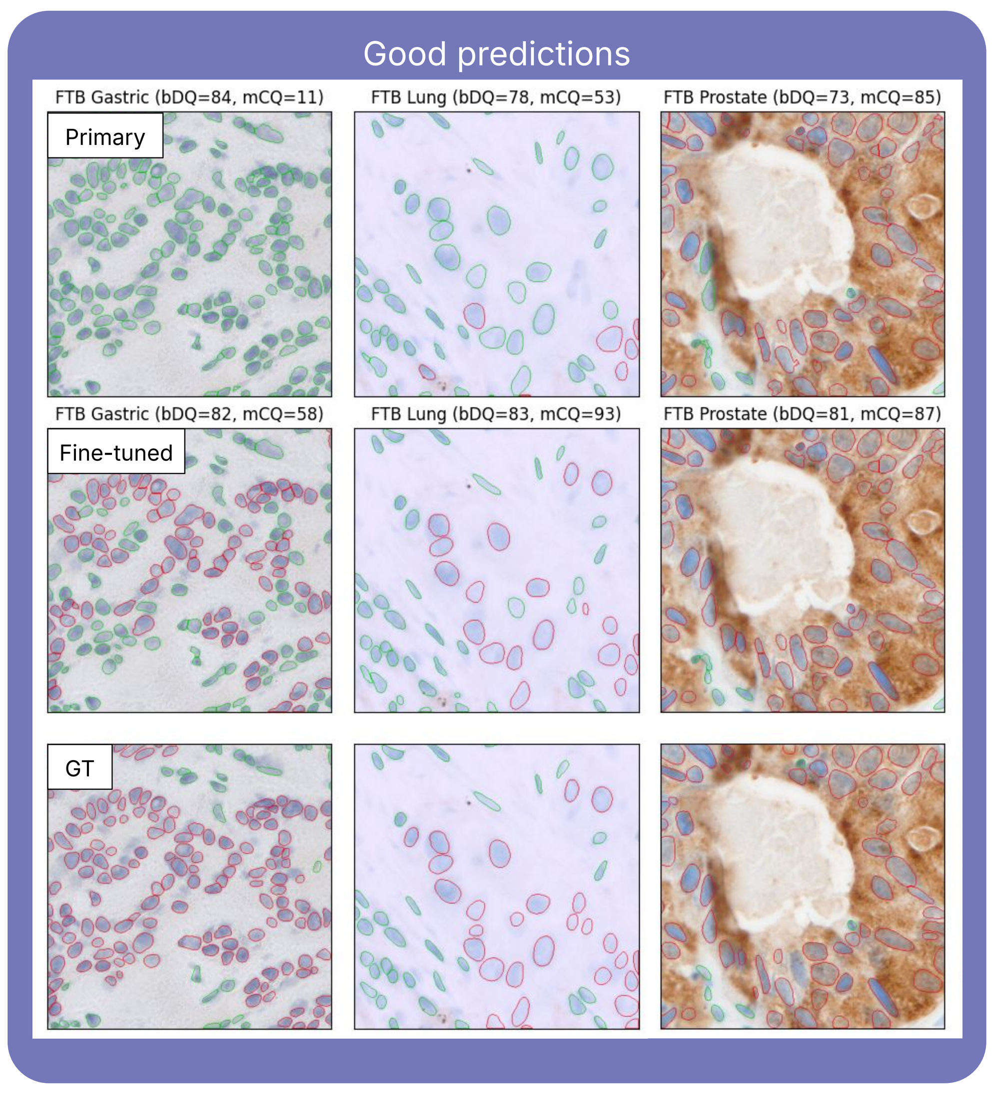


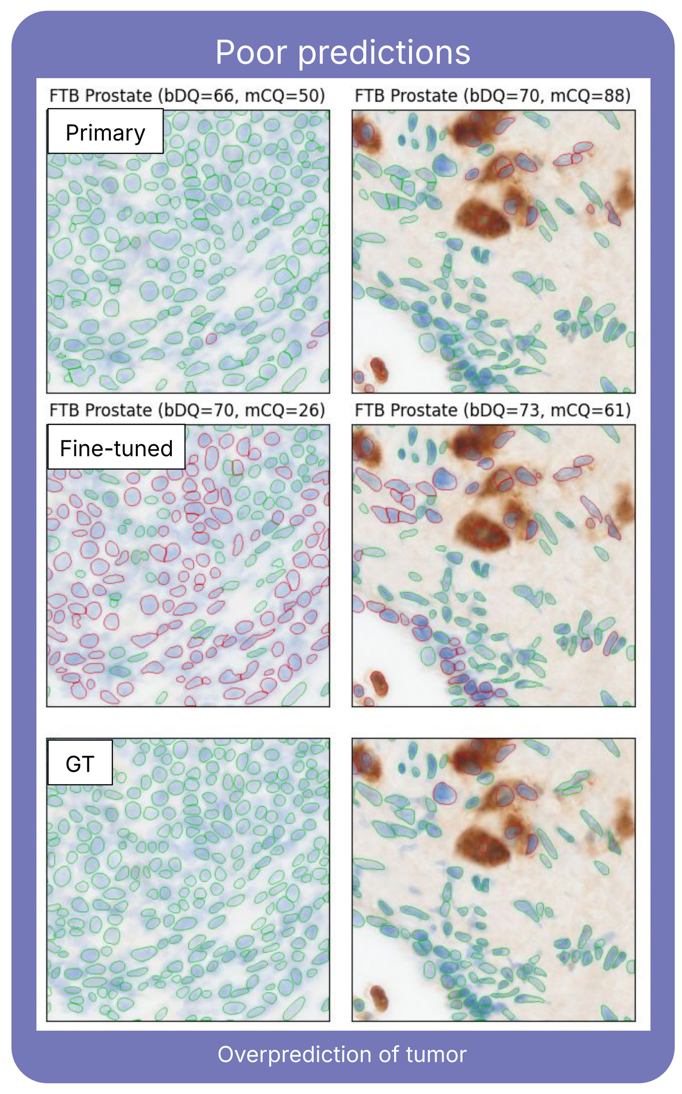
Tumor Other

Results: Fine-tuning

Source	Indication	Binary	Multiclass			
		DQ	CQ	DQ	SQ	PQ
FTB	Gastric	75 (+0)	77 (+9)	58 (+7)	79 (-1)	46 (+5)
	Lung	77 (+2)	63 (+2)	49 (+3)	79 (+1)	38 (+3)
	Prostate	74 (+1)	70 (+0)	52 (+1)	77 (+0)	40 (+1)

Fine-tuning metrics averaged over the validation sets of three cohort-stratified cross-validation folds, and a comparison with each corresponding metric of the primary model.





Discussion

The primary model generalizes better to the in-domain (breast HER2) cohort than those out-of-domain. Cell classification is the most domain-sensitive component as it shows the steepest performance drop at primary training and the greatest recovery after fine-tuning. Notably, joint fine-tuning for the three out-of-domain cohorts consistently improved or retained performance, supporting the feasibility of a truly tumor-agnostic model.

Conclusions

This prototype is the first step in the development of a tumor-agnostic AI-based IHC analysis pipeline. Making sure that these models can generalize to different tumor morphologies and IHC markers is crucial for future diagnostic applications.

References

1. N. A. Koohbanani, "NuClick: A Deep Learning Framework for Interactive Segmentation of Microscopy Images," *Medical Image Analysis*, Oct. 2020.
2. F. Hörst et al., "CellViT: Vision Transformers for precise cell segmentation and classification," *Medical Image Analysis*, May 2024.