STARC-9: A Large-scale Dataset for Multi-Class Tissue Classification for CRC Histopathology

Barathi Subramanian, Rathinaraja Jeyaraj, Mitchell Nevin Peterson, Terry Guo, Nigam Shah, Curtis Langlotz, Andrew Y. Ng, Jeanne Shen

Paper : https://arxiv.org/abs/2511.00383

Dataset : https://huggingface.co/datasets/Path2AI/STARC-9

Project code: https://github.com/Path2AI/STARC-9
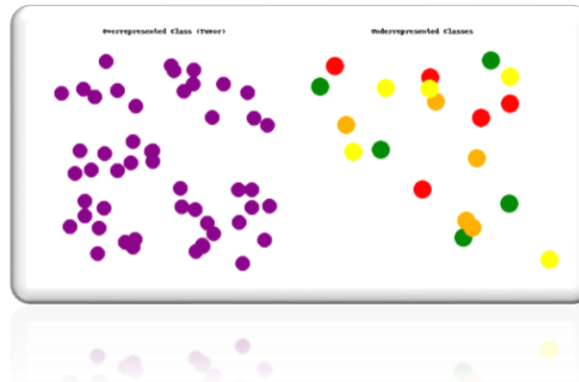
Core Challenges in Computational Pathology

🧩 Accurate tissue classification is the foundation for many AI diagnostics, but...

Why do most models fail to generalize? The data bottleneck

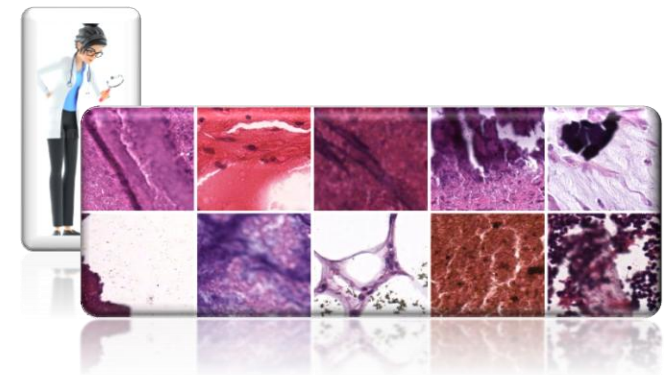**Problem 1:** Limited Availability of Public Datasets

Problem 2: Limited Morphologic Diversity/Class Imbalance

Problem 3: Low Quality or Mislabeled Tiles Need for Manual Curation

Objective

To create a dataset that is <mark>balanced</mark>, <mark>morphologically diverse</mark>, and <mark>quality-verified</mark>, enabling machine learning-model-ready training
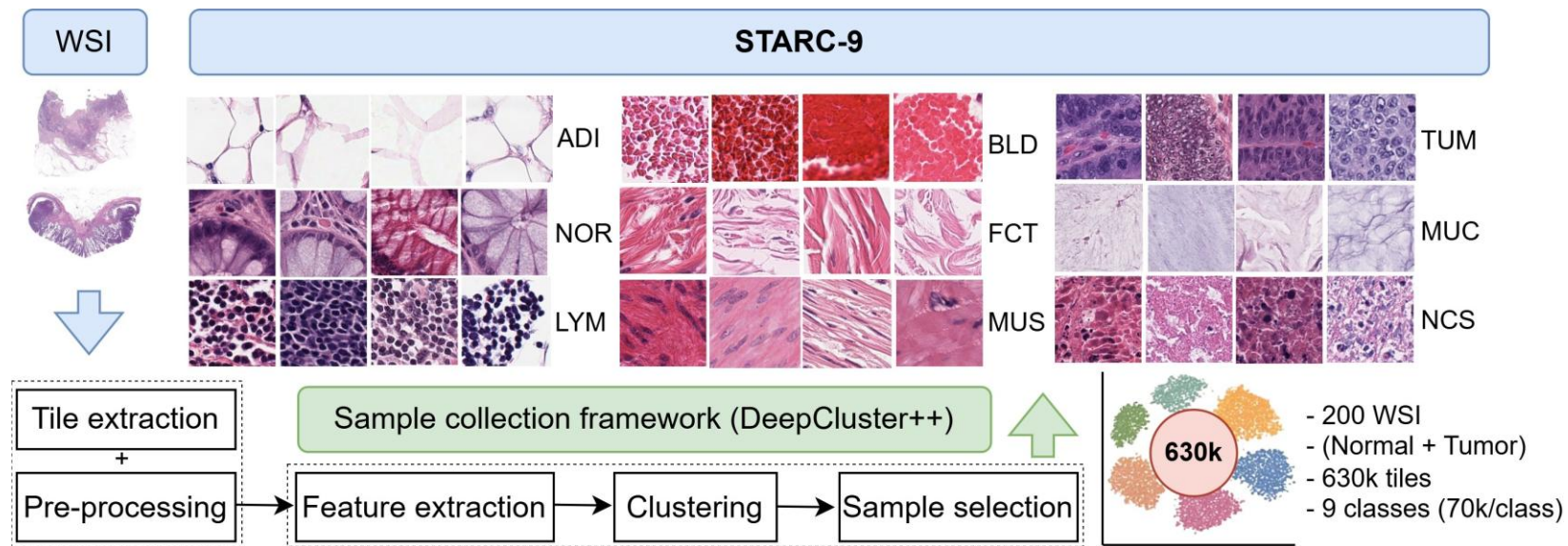

Figure 1: Overview of STARC-9 large-scale dataset generation. Adipose tissue (ADI), lymphoid tissue (LYM), muscle (MUS), fibroconnective tissue (FCT), mucin (MUC), Necrosis (NCS), blood (BLD), tumor (TUM), and normal mucosa (NOR)

DeepCluster++: STARC-9 Dataset Construction Framework

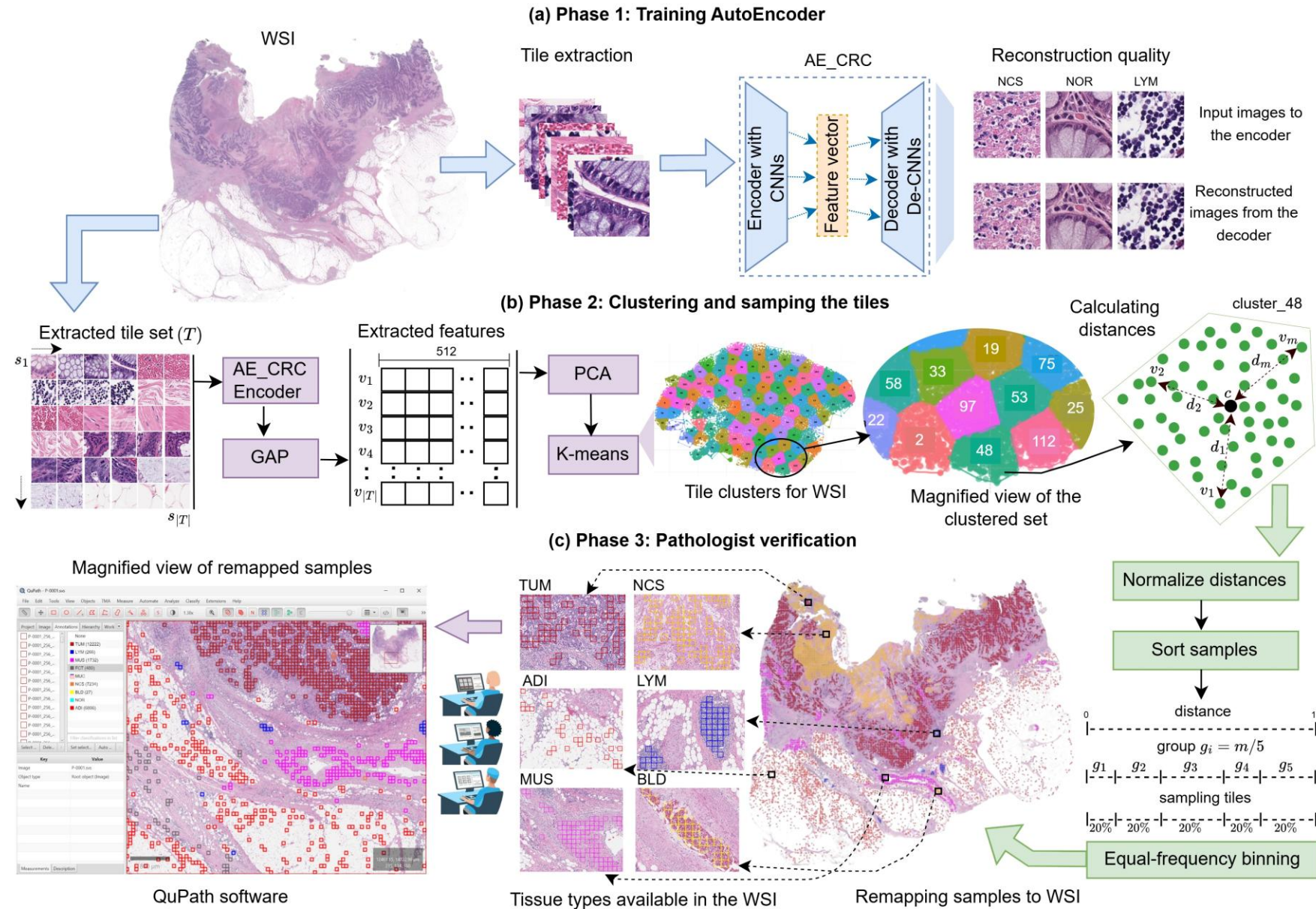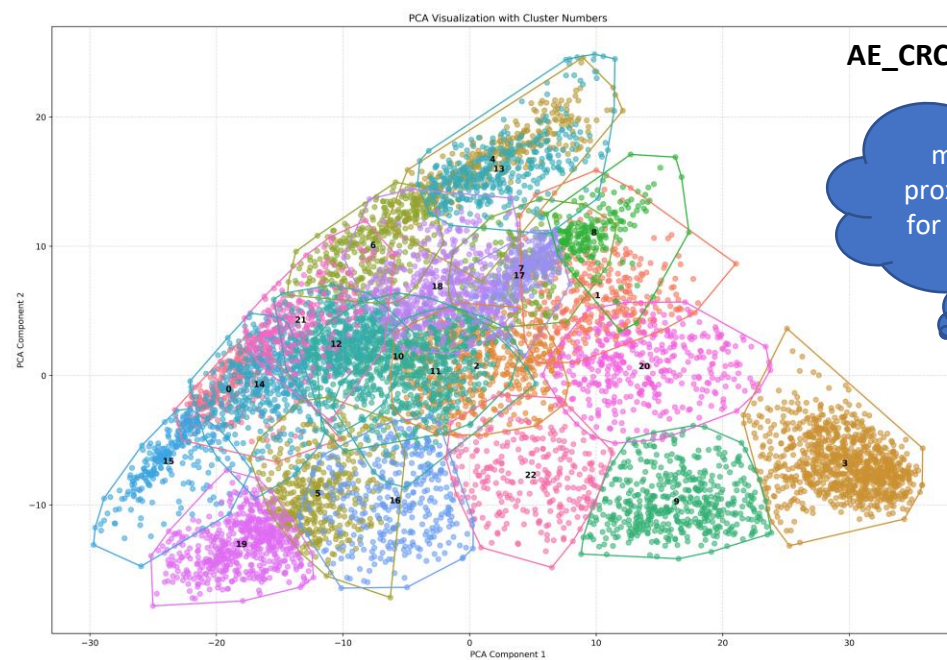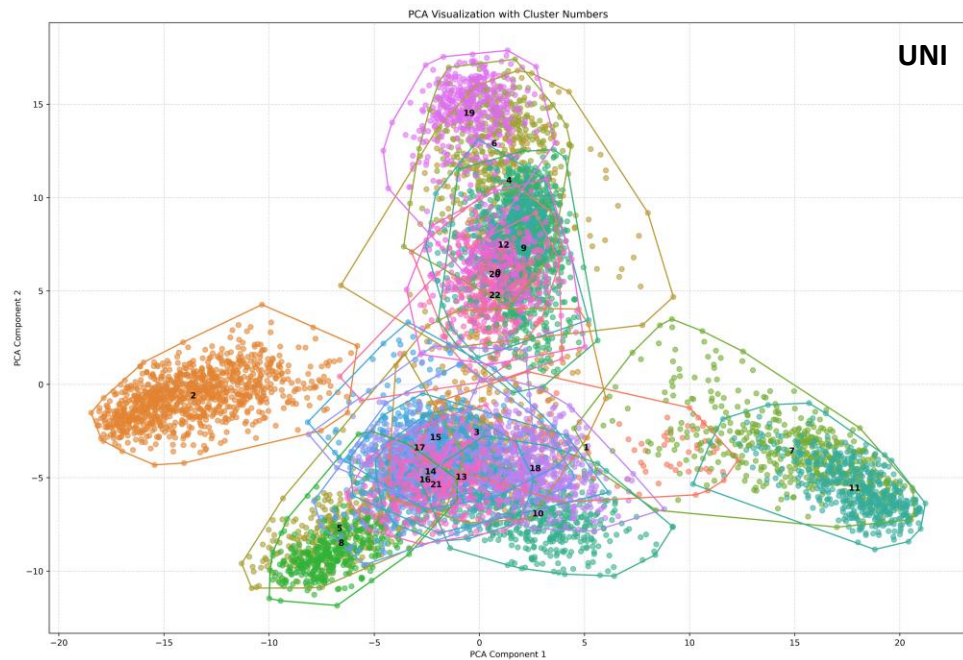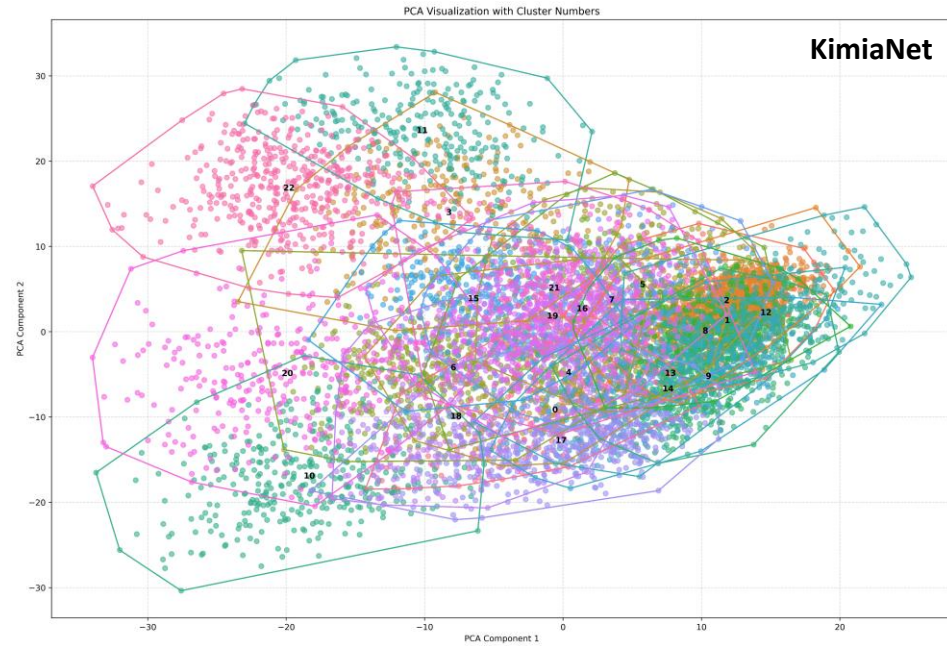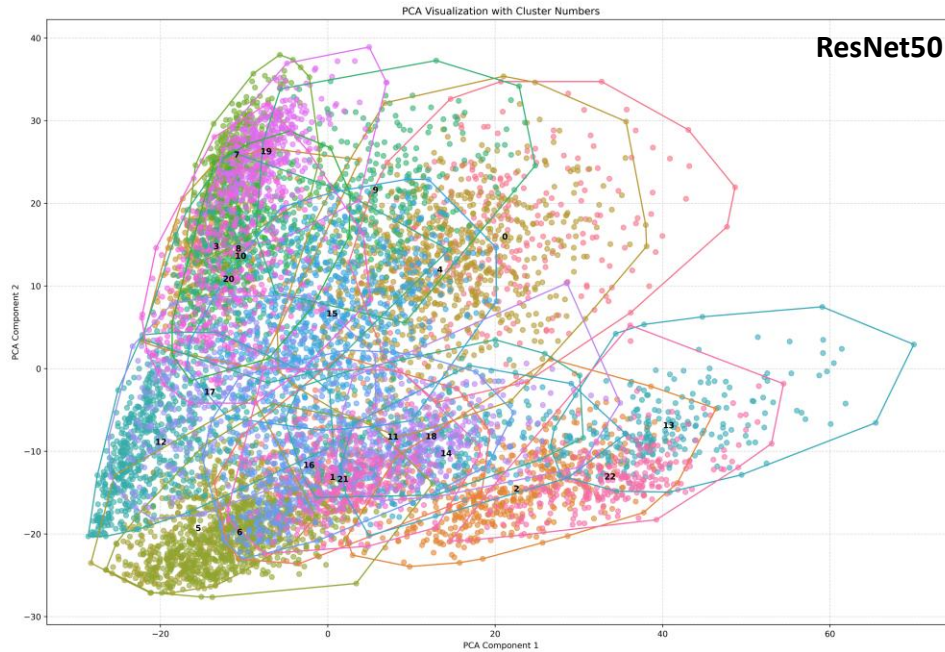**Figure 3**: DeepCluster++ framework (Phases 1 and 2) followed by pathologist verification (Phase 3)

Why AE_CRC? Feature Visualization for 9000 samples of nine tissue types

Preserves morphological proximity essential for diverse sample selection.

Results

Table 1: Multi-class classification performance of baseline, SOTA, pathology foundation, and custom models trained on HMU, NCT, and STARC-9 for seven common tissue types (ADI, LYM, MUS,MUC, NCS, TUM, NOR) on STANFORD-CRC-HE-VAL-LARGE dataset.

Model	Precision			Recall			F1-macro			Accuracy			Model size (M)
	NCT	HMU	STARC-9	NCT	HMU	STARC-9	NCT	HMU	STARC-9	NCT	HMU	STARC-9	
Baseline models													
ResNet50 [24]	84.08	87.81	98.92	62.59	85.71	98.64	63.17	86.00	98.78	62.59	85.71	98.64	24
EfficintNetB7 [25]	89.99	88.65	99.11	82.47	87.45	98.80	84.55	87.87	98.95	82.47	84.45	98.80	64
ViT-base [26]	92.71	91.57	98.49	84.25	90.29	98.09	87.30	90.87	98.28	**84.25**	90.29	98.09	86
SOTA models													
DeiT-B [27]	94.28	90.97	98.99	81.63	90.05	98.65	85.35	90.40	98.81	81.63	90.05	98.65	86
SwinTrans-b [28]	90.11	93.17	99.09	79.05	91.88	98.80	82.52	92.46	98.94	79.05	91.88	98.79	87
KimiaNet [20]	87.25	88.60	99.03	71.53	86.67	98.72	71.53	87.04	98.87	68.69	86.67	98.72	7
ConvNeXT-b [29]	91.95	92.09	99.01	82.82	91.07	98.36	85.56	91.50	98.68	82.82	91.07	98.36	88
Pathology foundation models													
CTransPath [30]	90.11	93.17	99.34	79.05	91.88	99.00	82.52	92.46	99.16	79.05	91.88	**99.00**	87
HiPT [31]	90.92	93.21	98.64	74.51	91.99	98.32	77.41	92.54	98.47	74.51	**91.99**	98.32	86
ProvGigPath [32]	89.43	91.47	98.74	74.18	90.60	98.37	78.40	90.92	98.55	74.18	90.60	98.37	305
PathDino [33]	92.93	91.19	98.67	77.35	89.64	98.37	81.71	90.22	98.51	77.35	89.64	98.37	22
CONCH [34]	91.53	91.41	98.56	75.69	90.02	98.19	78.08	90.52	98.37	75.69	90.02	98.19	87
UNI [35]	94.55	93.03	98.67	80.43	91.80	98.25	84.42	92.36	98.45	80.43	91.80	98.26	88
Virchow [36]	91.51	92.35	98.63	79.02	91.23	98.28	82.05	91.69	98.45	79.02	91.23	98.28	305
VIM4PATH [37]	92.66	92.66	98.53	75.41	91.50	98.27	79.10	92.01	98.40	75.41	91.50	98.29	86
Customized models (trained from scratch)													
CNN	83.97	78.45	98.10	64.21	68.10	97.81	68.12	66.39	97.93	64.21	68.10	97.81	3.9
Histo-ViT	86.17	76.45	96.88	69.48	67.16	96.32	72.01	67.77	96.52	69.48	67.16	96.32	86

Table 2: Multi-class classification performance of the best-performing models trained on HMU, NCT, and STARC-9 for seven common tissue types on the validation sets.

| Validation dataset	Precision			Recall			F1-macro			Accuracy		
	NCT	HMU	STARC-9	NCT	HMU	STARC-9	NCT	HMU	STARC-9	NCT	HMU	STARC-9
STANFORD-CRC-HE-VAL-SMALL	88.52	90.22	99.75	76.19	88.34	99.73	79.34	89.16	99.74	76.19	88.34	**99.73**
STANFORD-CRC-HE-VAL-LARGE	92.71	93.21	99.34	84.25	91.99	99.00	87.30	92.54	99.16	84.25	91.99	**99.00**
CURATED-TCGA-CRC-HE-VAL-20K	89.69	92.21	99.03	72.42	90.90	98.85	76.74	91.45	98.94	72.42	90.90	**98.85**
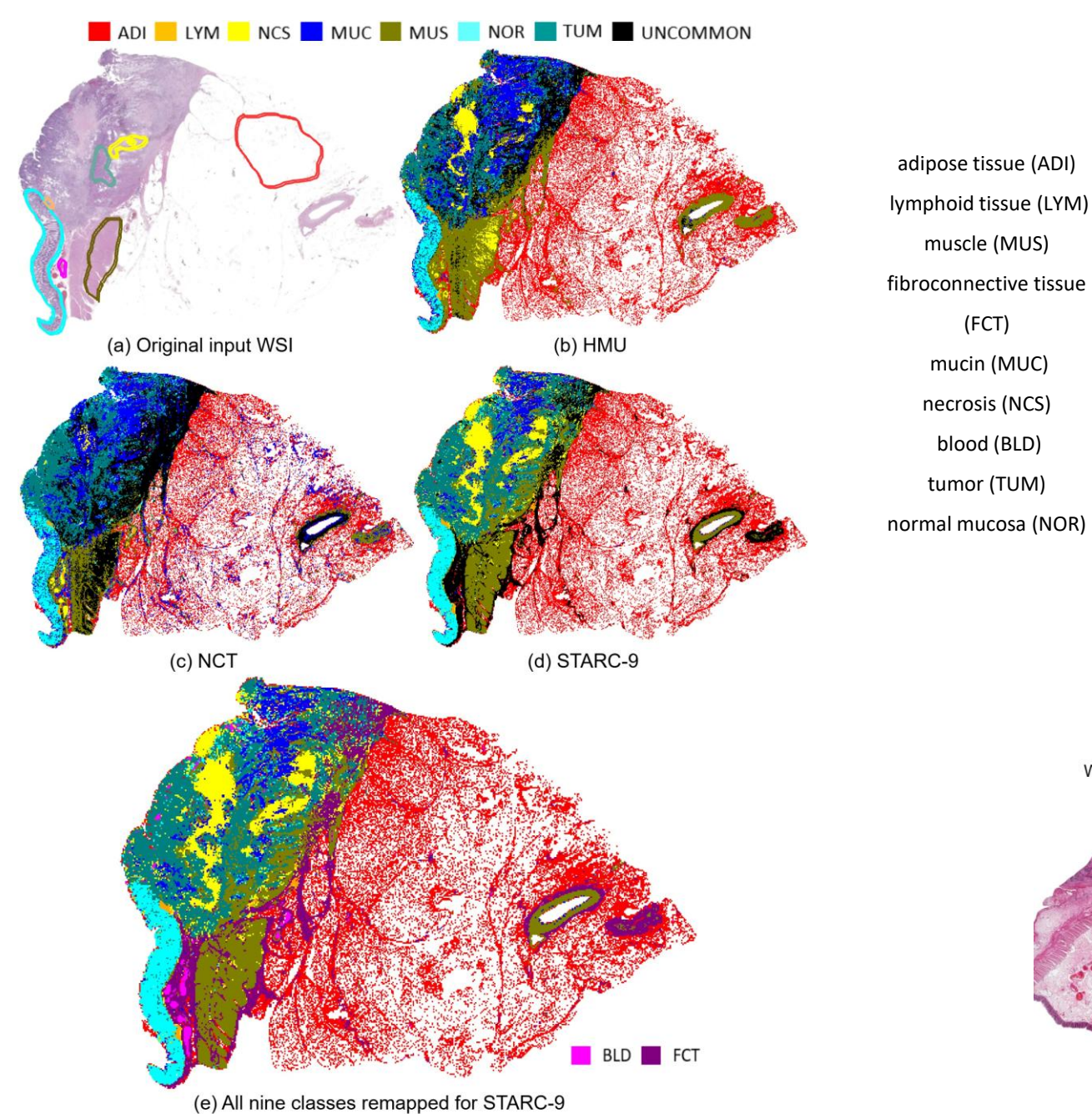
Figure 4: Tile-level predictions for common tissue types from models trained on different datasets, overlaid onto WSI.

Legend: ADI, LYM, NCS, MUC, MUS, NOR, TUM, UNCOMMON

adipose tissue (ADI)

lymphoid tissue (LYM)

muscle (MUS)

fibroconnective tissue (FCT)

mucin (MUC)

necrosis (NCS)

blood (BLD)

tumor (TUM)

normal mucosa (NOR)

(a) Original input WSI

(b) HMU

(c) NCT

(d) STARC-9

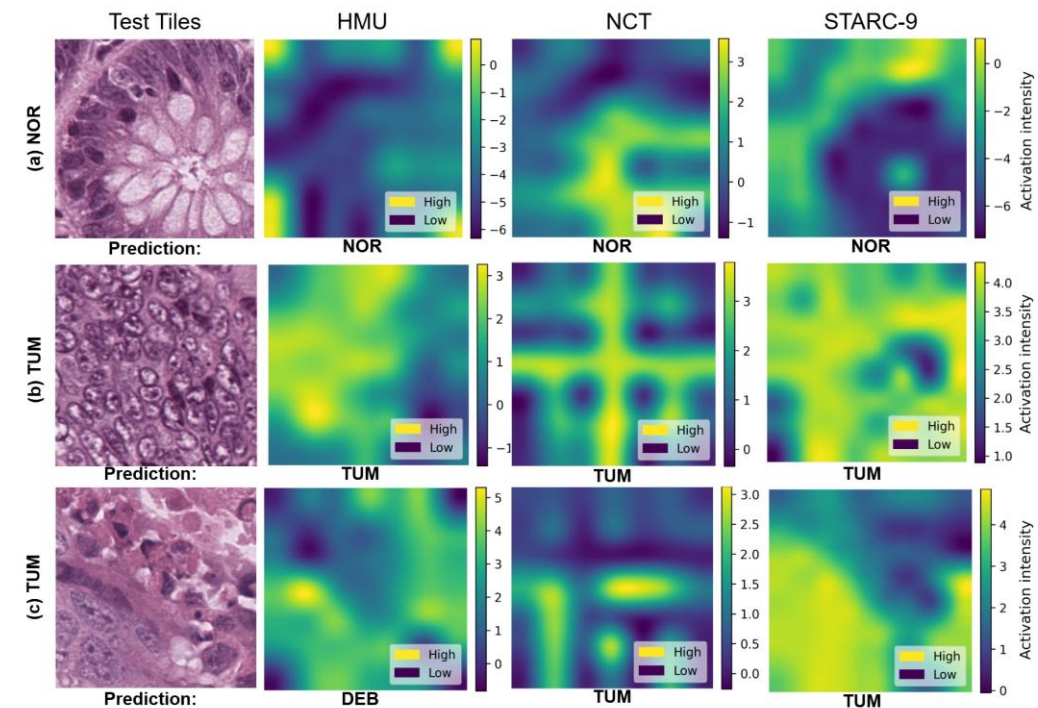(e) All nine classes remapped for STARC-9

BLD, FCT

Figure 5: Feature map visualizations for the best models trained on HMU, NCT and STARC-9
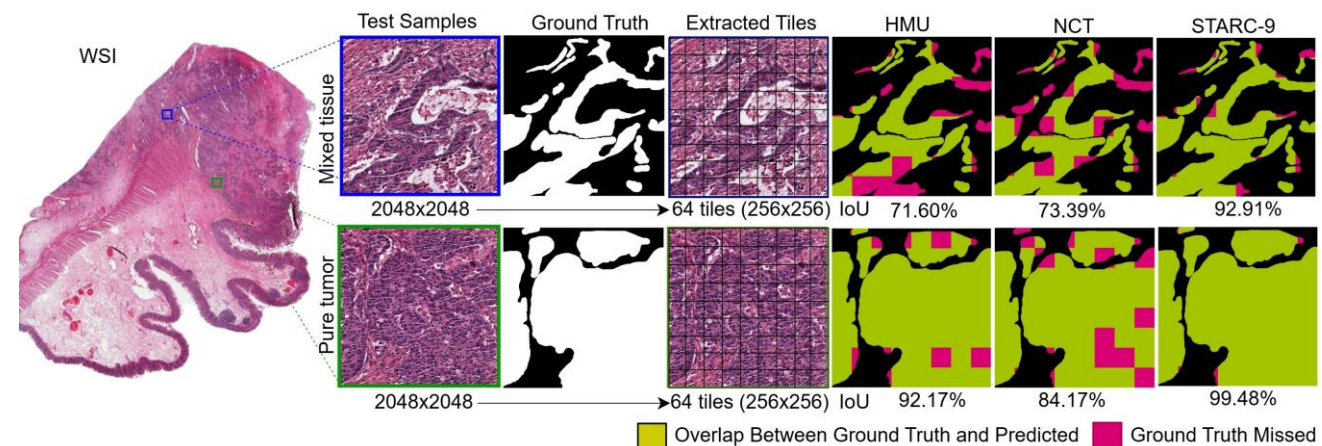

Figure 6: Tumor segmentation within 2048x2048 regions from a WSI from the CURATED-TCGA-CRC-HE-VAL dataset using tile-level classifiers trained on HMU, NCT, and STARC-9

Conclusions and Limitations

- We introduced a large-scale, quality-verified dataset for colorectal tissue classification.

- We proposed DeepCluster++, a scalable semi-automated sampling and curation pipeline.

- STARC-9-trained models show improved generalization across datasets.

Limitations:

➤ Dataset limited to colorectal slides

➤ Multi-institutional diversity still under expansion

Future work: We plan to extend STARC-9 to multi-organ datasets, integrate multimodal data such as genomics, and make the dataset publicly available.

THANK YOU FOR YOUR ATTENTION

Any questions?

Please reach out to us at: barathi.sn93@gmail.com, jrathinaraja@gmail.com

Scan me!