Mahmood Lab
AI for Pathology

Visual Language Pretrained Multiple Instance Zero-Shot Transfer for Histopathology Images

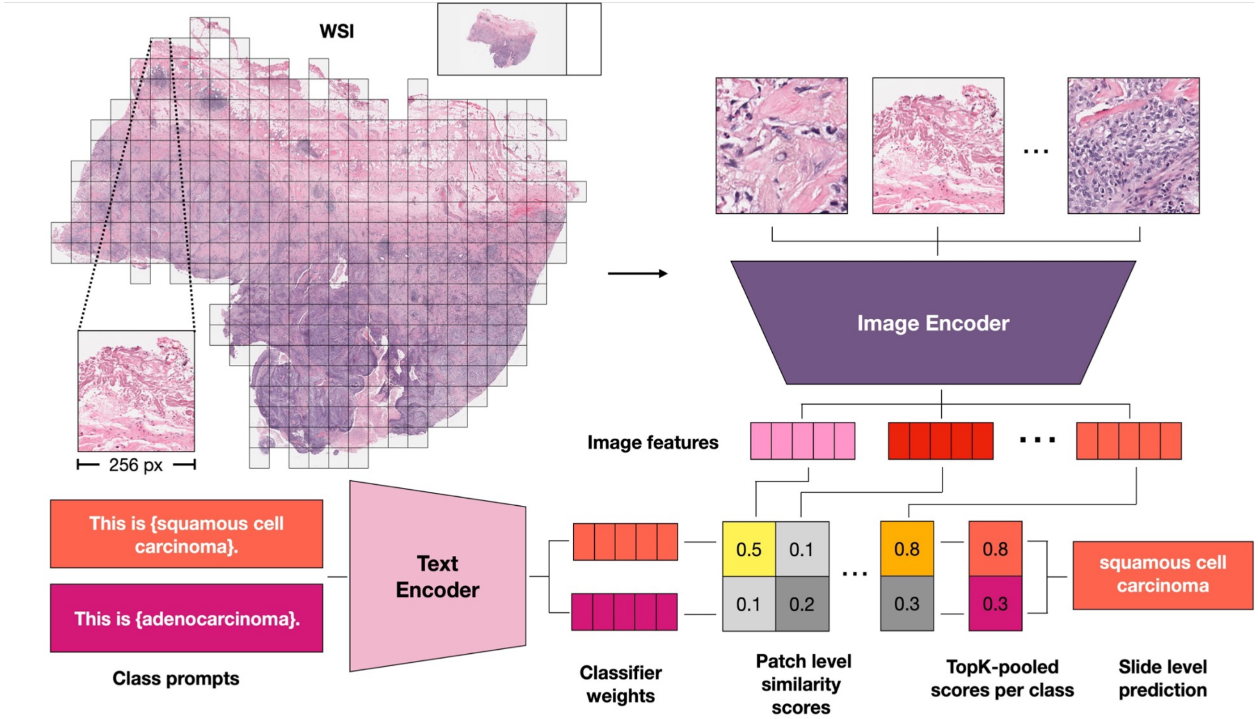
Ming Y. Lu[*,1,2,3], Bowen Chen[*,2,3], Andrew Zhang[1,2,3], Drew F. K. Williamson[2,3], Richard J. Chen[2,3], Tong Ding[2,3], Long Phi Le[2,3], Yung-Sung Chuang[1], Faisal Mahmood[2,3]

[1]MIT, [2]Harvard, [3]BWH
[*]Contributed Equally

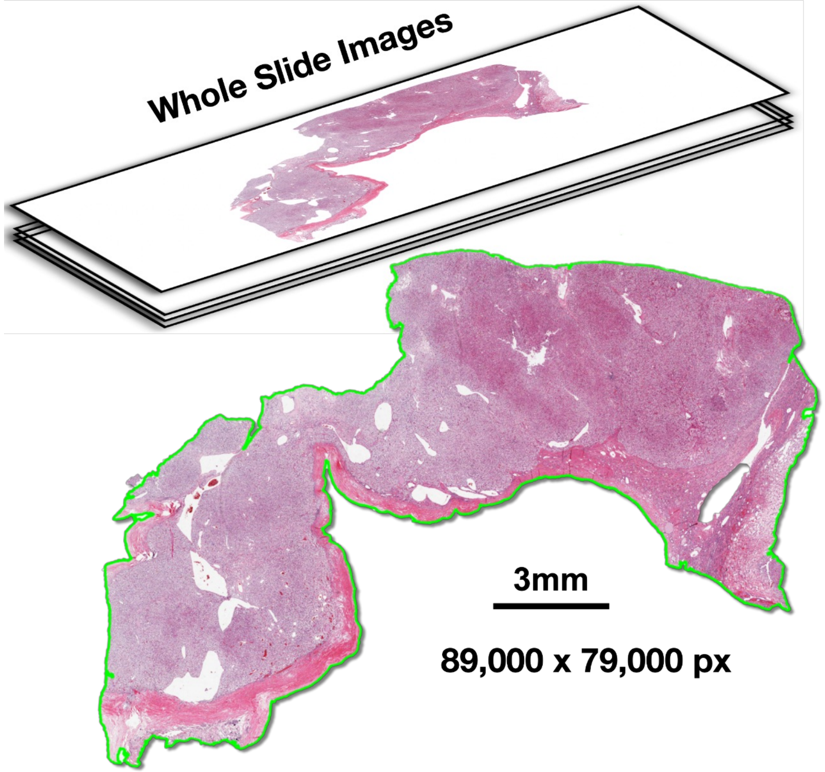
THU-AM-312

HARVARD MEDICAL SCHOOL   Mass General Brigham   BROAD INSTITUTE   Dana-Farber Cancer Institute   JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Overview


WSI
256 px
Image Encoder
Image features
This is {squamous cell carcinoma}.
This is {adenocarcinoma}.
Class prompts
Text Encoder
Classifier weights
0.5
0.1
0.1
0.2
Patch level similarity scores
0.8
0.3
0.8
0.3
TopK-pooled scores per class
squamous cell carcinoma
Slide level prediction

Current paradigm in computational pathology



Predict →

Cancer vs. Non-cancer?

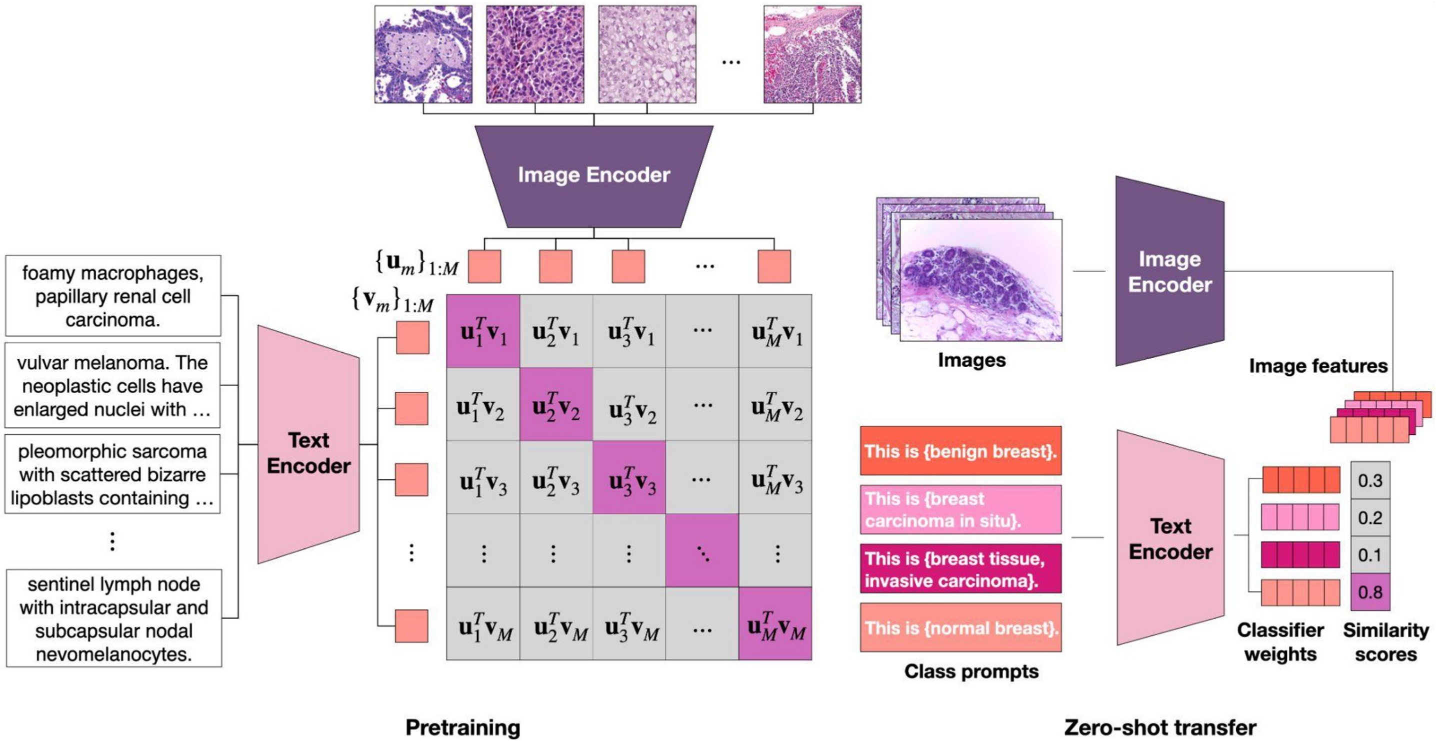
Metastatic vs. Primary?

Subtype A vs. Subtype B vs. Subtype C?

High-grade vs. Low-grade?

…

Contrastive image text pretraining (CLIP) in general VL


...
Image Encoder
$\{\mathbf{u}_m\}_{1:M}$
$\{\mathbf{v}_m\}_{1:M}$
foamy macrophages, papillary renal cell carcinoma.
vulvar melanoma. The neoplastic cells have enlarged nuclei with ...
pleomorphic sarcoma with scattered bizarre lipoblasts containing ...
⋮
sentinel lymph node with intracapsular and subcapsular nodal nevomelanocytes.
Text Encoder
$\mathbf{u}_1^T\mathbf{v}_1$ $\mathbf{u}_2^T\mathbf{v}_1$ $\mathbf{u}_3^T\mathbf{v}_1$ ... $\mathbf{u}_M^T\mathbf{v}_1$
$\mathbf{u}_1^T\mathbf{v}_2$ $\mathbf{u}_2^T\mathbf{v}_2$ $\mathbf{u}_3^T\mathbf{v}_2$ ... $\mathbf{u}_M^T\mathbf{v}_2$
$\mathbf{u}_1^T\mathbf{v}_3$ $\mathbf{u}_2^T\mathbf{v}_3$ $\mathbf{u}_3^T\mathbf{v}_3$ ... $\mathbf{u}_M^T\mathbf{v}_3$
$\mathbf{u}_1^T\mathbf{v}_M$ $\mathbf{u}_2^T\mathbf{v}_M$ $\mathbf{u}_3^T\mathbf{v}_M$ ... $\mathbf{u}_M^T\mathbf{v}_M$
Pretraining
Images
Image Encoder
Image features
This is {benign breast}.
This is {breast carcinoma in situ}.
This is {breast tissue, invasive carcinoma}.
This is {normal breast}.
Class prompts
Text Encoder
Classifier weights
0.3
0.2
0.1
0.8
Similarity scores
Zero-shot transfer

Radford *et al.*, Learning Transferable Visual Models From Natural Language Supervision. PMLR 2021

Challenges for computational pathology

- Lack of paired data
  - Need to curate large-scale domain-specific dataset
- Gigapixel images
  - Need to generalize downstream functionalities of aligned encoders for extremely large images

Dataset	**Number of samples**
MS-COCO (Lin *et al.*)	>200k labeled
YFCC100M (Thomee *et al.*)	99.2M
LAION 400M (Schuhmann *et al.*)	400M
CLIP (Radford *et al.*)	400M
ALIGN (Jia *et al.*)	1.8B
LiT (Zhai *et al.*)	4B
LAION 5B (Schuhmann *et al.*)	5.85B
ARCH (Gamper *et al.*)	7.6k

Curating paired image-text histopathology dataset

Scrape

- Scraping images-caption pairs from publicly-available pathology education resources
- Combine with existing available image-caption dataset (ARCH)

Filter

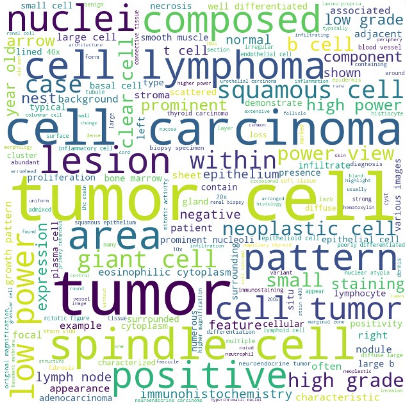
- Keep histopathology microscopy images
- Remove:
    - Gross
    - Cytology
    - X-ray/CT
    - EM
    - Fluorescent
    - Schematics

Clean

- Crop multipanel figures (and separate captions accordingly)
- Remove courtesy and acknowledgements
- Remove figures IDs

Result

- 33,480 image-text pairs
- Largest histopathology image-text dataset at the time of study



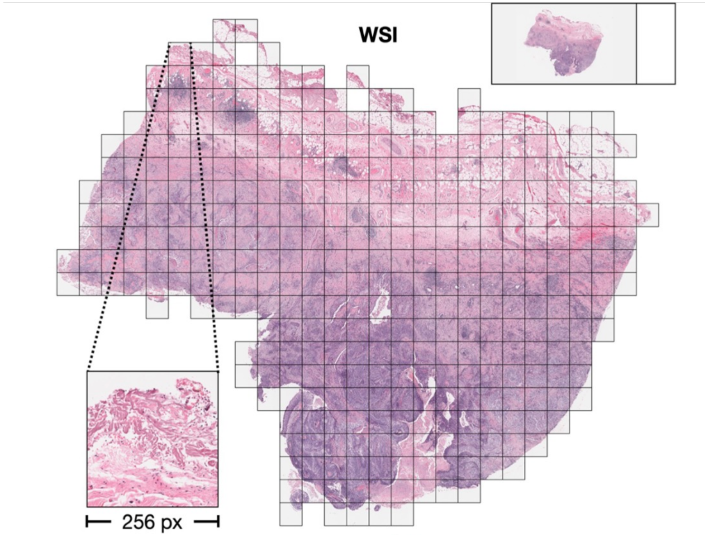
Leveraging pretrained unimodal encoders

Vision Encoder	Pretraining Data	Domain
CTP	Histopathology image patches	In-domain (histopathology)
ViT-S	Histopathology image patches	In-domain (histopathology)
ViT-S	ImageNet supervised	Out-of-domain (general vision)

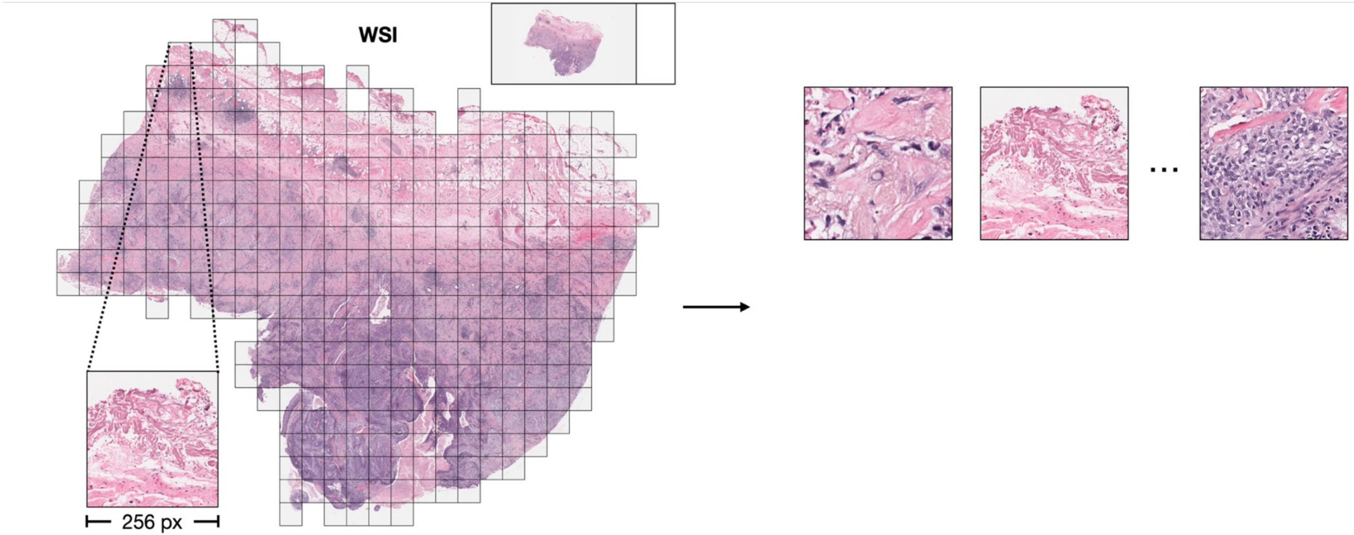
Text Encoder	Pretraining Data	Domain
HistPathGPT	Histopathology-relevant corpora (e.g. surgical reports)	In-domain (histopathology)
BioClinicalBert	MIMIC III	Out-of-domain (general biomedical text)
PubmedBert	PubMed abstracts	Out-of-domain (general biomedical text)

Wang *et al.,* Transformer-based unsupervised contrastive learning for histopathological image classification. *Medical Image Analysis* 2022
Chen *et al.*, An empirical study of training self-supervised vision transformers. CVPR 2021

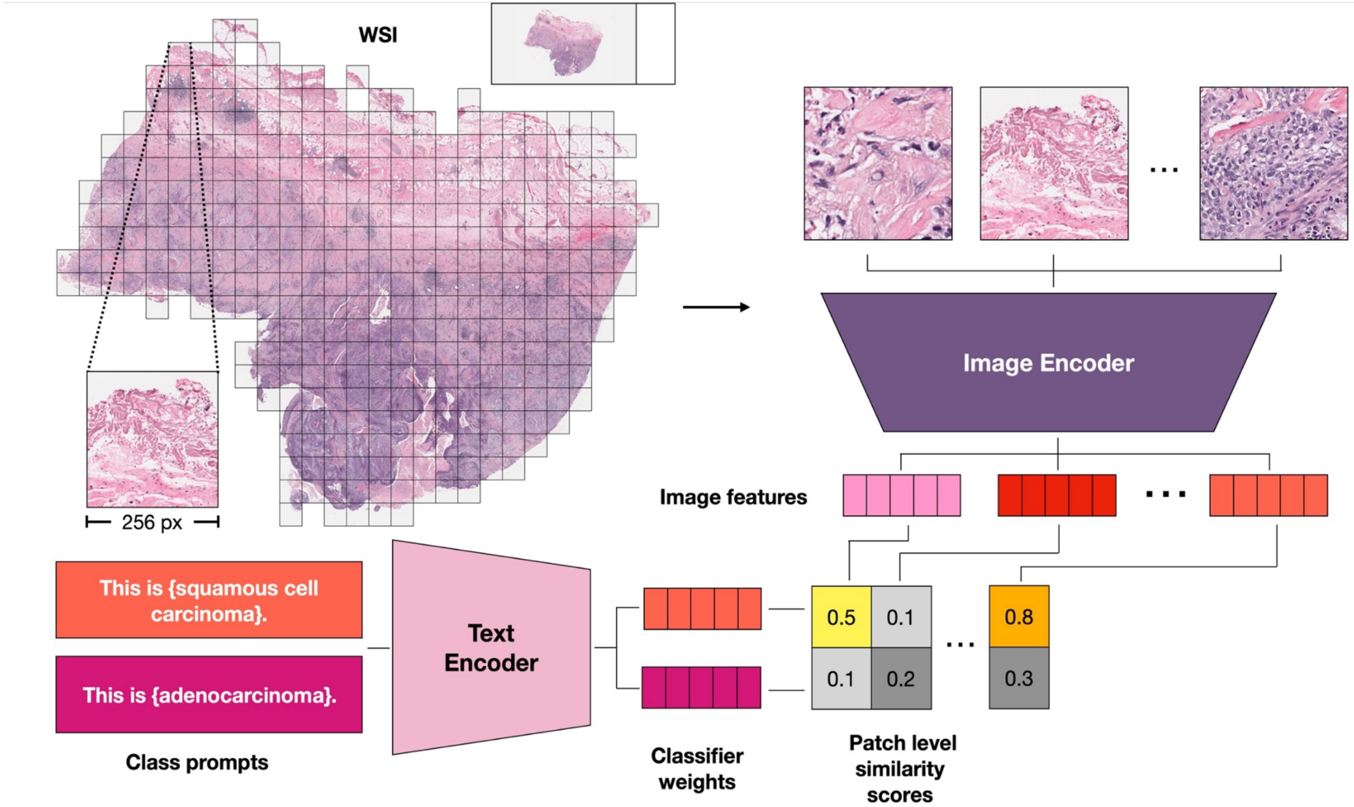
MI-Zero scales zero-shot transfer to gigapixel images



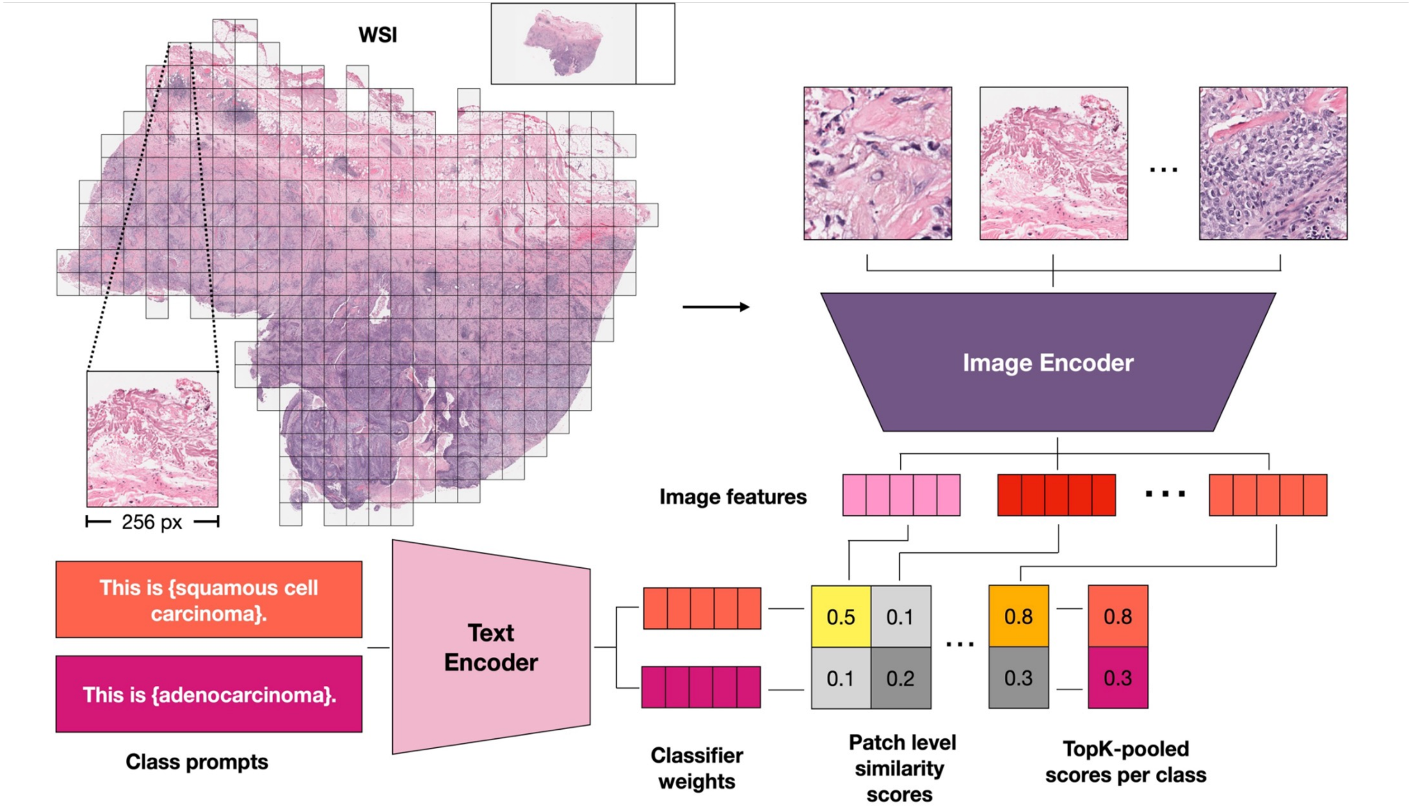
MI-Zero scales zero-shot transfer to gigapixel images



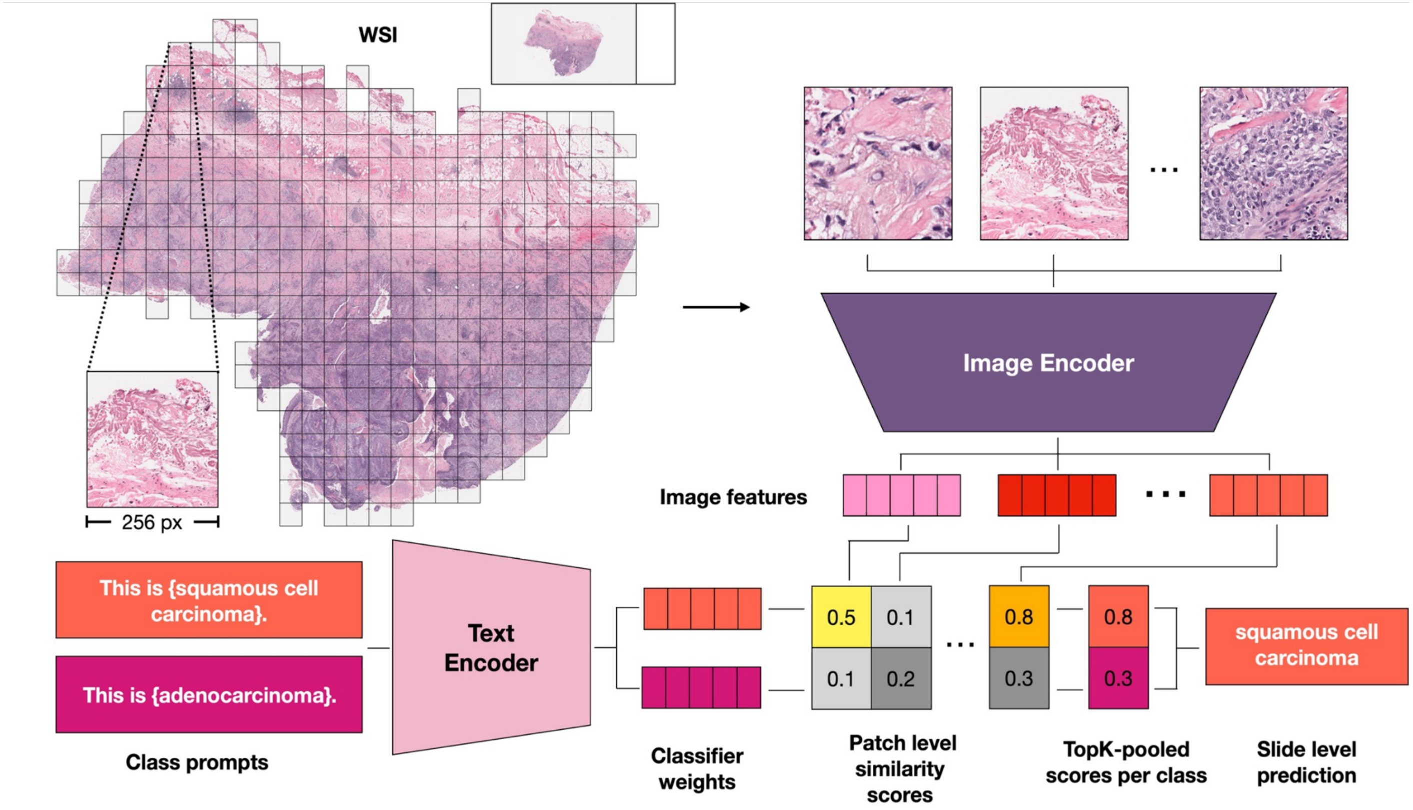
MI-Zero scales zero-shot transfer to gigapixel images



MI-Zero scales zero-shot transfer to gigapixel images


WSI
256 px
Image Encoder
Image features
This is {squamous cell carcinoma}.
This is {adenocarcinoma}.
Class prompts
Text Encoder
Classifier weights
0.5
0.1
0.1
0.2
Patch level similarity scores
0.8
0.3
0.8
0.3
TopK-pooled scores per class

MI-Zero scales zero-shot transfer to gigapixel images



Downstream evaluation for zero-shot transfer

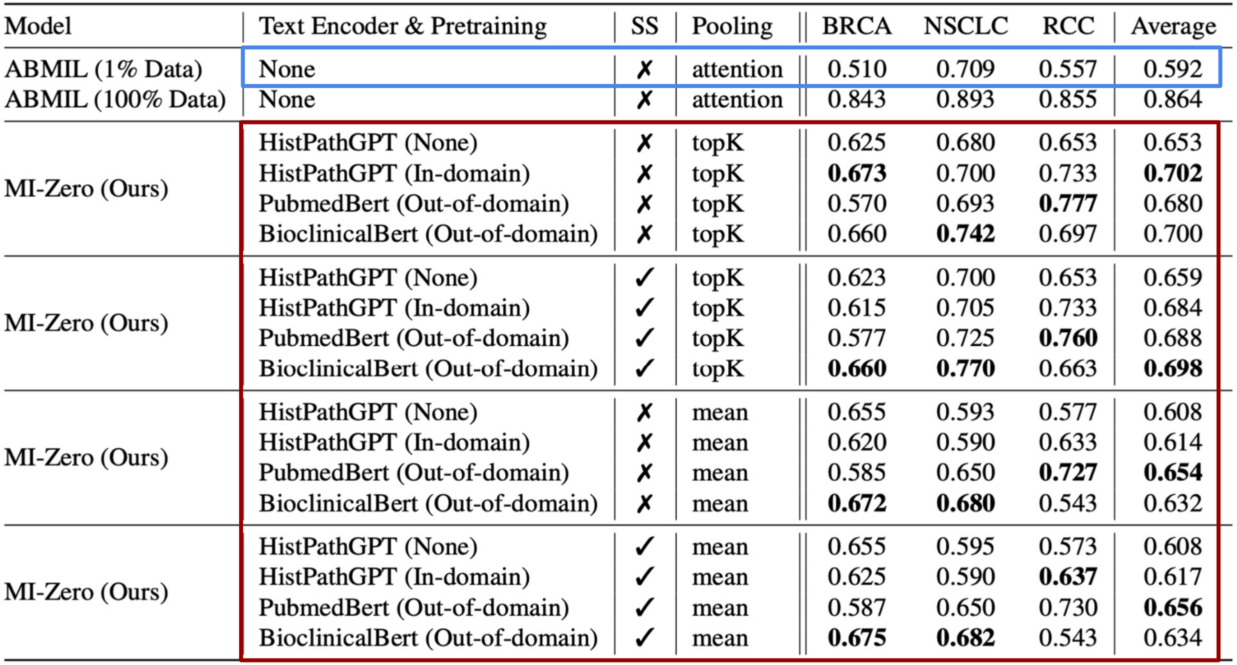
Tasks (in-house WSIs):

- BRCA subtyping
- NSCLC subtyping
- RCC subtyping

Evaluation method:

- 256 × 256px patches at 20× equivalent magnification
- Curate a list of text prompts suggested by a pathologist
- Sample 50 subsets of prompts and compute balanced accuracy for each iteration
- Compute median balanced accuracy over the 50 iterations

Slide-level zero-shot transfer: ours vs ABMIL baseline

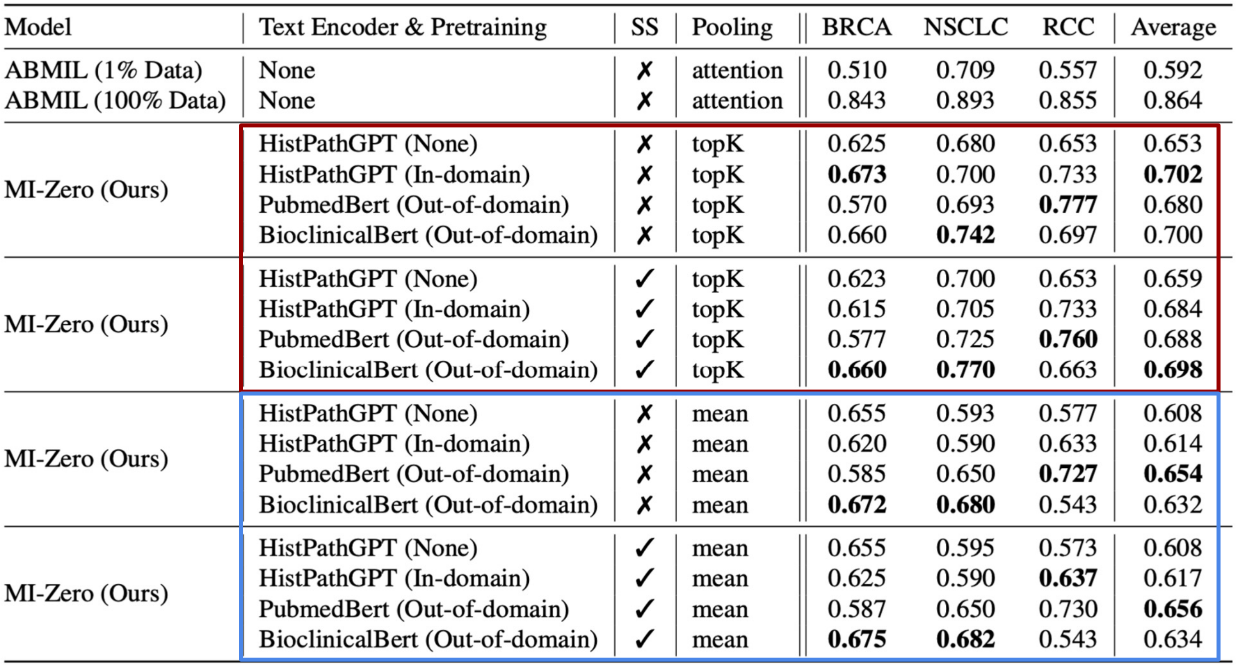


Model	Text Encoder & Pretraining	SS	Pooling	BRCA	NSCLC	RCC	Average
ABMIL (1% Data)	None	✗	attention	0.510	0.709	0.557	0.592
ABMIL (100% Data)	None	✗	attention	0.843	0.893	0.855	0.864
MI-Zero (Ours)	HistPathGPT (None)	✗	topK	0.625	0.680	0.653	0.653
	HistPathGPT (In-domain)	✗	topK	**0.673**	0.700	0.733	**0.702**
	PubmedBert (Out-of-domain)	✗	topK	0.570	0.693	**0.777**	0.680
	BioclinicalBert (Out-of-domain)	✗	topK	0.660	**0.742**	0.697	0.700
MI-Zero (Ours)	HistPathGPT (None)	✓	topK	0.623	0.700	0.653	0.659
	HistPathGPT (In-domain)	✓	topK	0.615	0.705	0.733	0.684
	PubmedBert (Out-of-domain)	✓	topK	0.577	0.725	**0.760**	0.688
	BioclinicalBert (Out-of-domain)	✓	topK	**0.660**	**0.770**	0.663	**0.698**
MI-Zero (Ours)	HistPathGPT (None)	✗	mean	0.655	0.593	0.577	0.608
	HistPathGPT (In-domain)	✗	mean	0.620	0.590	0.633	0.614
	PubmedBert (Out-of-domain)	✗	mean	0.585	0.650	**0.727**	**0.654**
	BioclinicalBert (Out-of-domain)	✗	mean	**0.672**	**0.680**	0.543	0.632
MI-Zero (Ours)	HistPathGPT (None)	✓	mean	0.655	0.595	0.573	0.608
	HistPathGPT (In-domain)	✓	mean	0.625	0.590	**0.637**	0.617
	PubmedBert (Out-of-domain)	✓	mean	0.587	0.650	0.730	**0.656**
	BioclinicalBert (Out-of-domain)	✓	mean	**0.675**	**0.682**	0.543	0.634

1% ABMIL

Ours

Slide-level zero-shot transfer: pooling method

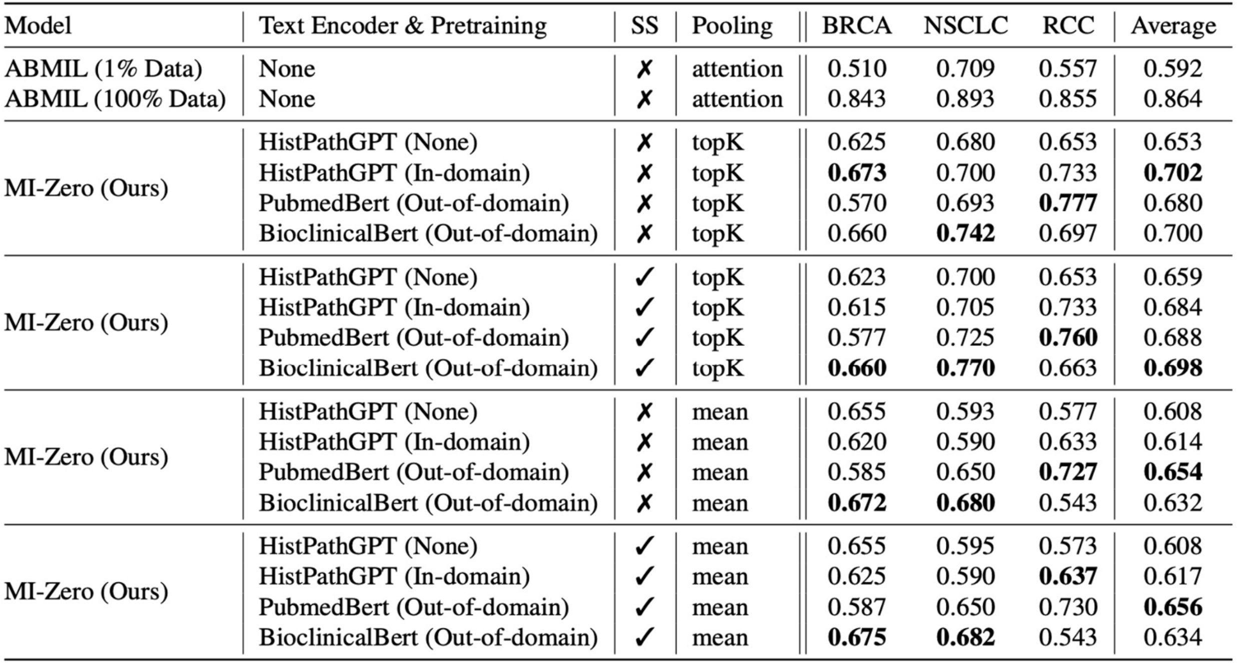


Model	Text Encoder & Pretraining	SS	Pooling	BRCA	NSCLC	RCC	Average
ABMIL (1% Data)	None	✗	attention	0.510	0.709	0.557	0.592
ABMIL (100% Data)	None	✗	attention	0.843	0.893	0.855	0.864
MI-Zero (Ours)	HistPathGPT (None)	✗	topK	0.625	0.680	0.653	0.653
	HistPathGPT (In-domain)	✗	topK	**0.673**	0.700	0.733	**0.702**
	PubmedBert (Out-of-domain)	✗	topK	0.570	0.693	**0.777**	0.680
	BioclinicalBert (Out-of-domain)	✗	topK	0.660	**0.742**	0.697	0.700
MI-Zero (Ours)	HistPathGPT (None)	✓	topK	0.623	0.700	0.653	0.659
	HistPathGPT (In-domain)	✓	topK	0.615	0.705	0.733	0.684
	PubmedBert (Out-of-domain)	✓	topK	0.577	0.725	**0.760**	0.688
	BioclinicalBert (Out-of-domain)	✓	topK	**0.660**	**0.770**	0.663	**0.698**
MI-Zero (Ours)	HistPathGPT (None)	✗	mean	0.655	0.593	0.577	0.608
	HistPathGPT (In-domain)	✗	mean	0.620	0.590	0.633	0.614
	PubmedBert (Out-of-domain)	✗	mean	0.585	0.650	**0.727**	**0.654**
	BioclinicalBert (Out-of-domain)	✗	mean	**0.672**	**0.680**	0.543	0.632
MI-Zero (Ours)	HistPathGPT (None)	✓	mean	0.655	0.595	0.573	0.608
	HistPathGPT (In-domain)	✓	mean	0.625	0.590	**0.637**	0.617
	PubmedBert (Out-of-domain)	✓	mean	0.587	0.650	0.730	**0.656**
	BioclinicalBert (Out-of-domain)	✓	mean	**0.675**	**0.682**	0.543	0.634

TopK pooling

Mean pooling

Slide-level zero-shot transfer: spatial smoothing

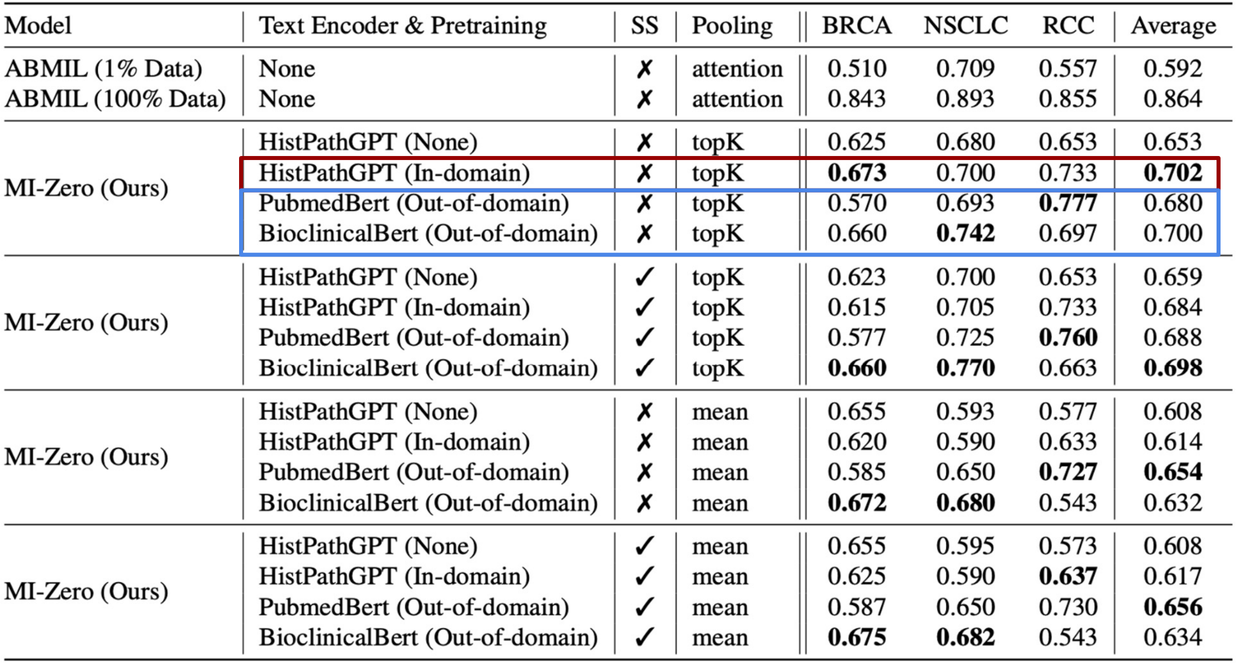


Model	Text Encoder & Pretraining	SS	Pooling	BRCA	NSCLC	RCC	Average
ABMIL (1% Data)	None	✗	attention	0.510	0.709	0.557	0.592
ABMIL (100% Data)	None	✗	attention	0.843	0.893	0.855	0.864
MI-Zero (Ours)	HistPathGPT (None)	✗	topK	0.625	0.680	0.653	0.653
	HistPathGPT (In-domain)	✗	topK	**0.673**	0.700	0.733	**0.702**
	PubmedBert (Out-of-domain)	✗	topK	0.570	0.693	**0.777**	0.680
	BioclinicalBert (Out-of-domain)	✗	topK	0.660	**0.742**	0.697	0.700
MI-Zero (Ours)	HistPathGPT (None)	✓	topK	0.623	0.700	0.653	0.659
	HistPathGPT (In-domain)	✓	topK	0.615	0.705	0.733	0.684
	PubmedBert (Out-of-domain)	✓	topK	0.577	0.725	**0.760**	0.688
	BioclinicalBert (Out-of-domain)	✓	topK	**0.660**	**0.770**	0.663	**0.698**
MI-Zero (Ours)	HistPathGPT (None)	✗	mean	0.655	0.593	0.577	0.608
	HistPathGPT (In-domain)	✗	mean	0.620	0.590	0.633	0.614
	PubmedBert (Out-of-domain)	✗	mean	0.585	0.650	**0.727**	**0.654**
	BioclinicalBert (Out-of-domain)	✗	mean	**0.672**	**0.680**	0.543	0.632
MI-Zero (Ours)	HistPathGPT (None)	✓	mean	0.655	0.595	0.573	0.608
	HistPathGPT (In-domain)	✓	mean	0.625	0.590	**0.637**	0.617
	PubmedBert (Out-of-domain)	✓	mean	0.587	0.650	0.730	**0.656**
	BioclinicalBert (Out-of-domain)	✓	mean	**0.675**	**0.682**	0.543	0.634

No spatial smoothing

Spatial smoothing

Slide-level zero-shot transfer: text pretraining



Model	Text Encoder & Pretraining	SS	Pooling	BRCA	NSCLC	RCC	Average
ABMIL (1% Data)	None	✗	attention	0.510	0.709	0.557	0.592
ABMIL (100% Data)	None	✗	attention	0.843	0.893	0.855	0.864
MI-Zero (Ours)	HistPathGPT (None)	✗	topK	0.625	0.680	0.653	0.653
	HistPathGPT (In-domain)	✗	topK	**0.673**	0.700	0.733	**0.702**
	PubmedBert (Out-of-domain)	✗	topK	0.570	0.693	**0.777**	0.680
	BioclinicalBert (Out-of-domain)	✗	topK	0.660	**0.742**	0.697	0.700
MI-Zero (Ours)	HistPathGPT (None)	✓	topK	0.623	0.700	0.653	0.659
	HistPathGPT (In-domain)	✓	topK	0.615	0.705	0.733	0.684
	PubmedBert (Out-of-domain)	✓	topK	0.577	0.725	**0.760**	0.688
	BioclinicalBert (Out-of-domain)	✓	topK	**0.660**	**0.770**	0.663	**0.698**
MI-Zero (Ours)	HistPathGPT (None)	✗	mean	0.655	0.593	0.577	0.608
	HistPathGPT (In-domain)	✗	mean	0.620	0.590	0.633	0.614
	PubmedBert (Out-of-domain)	✗	mean	0.585	0.650	**0.727**	**0.654**
	BioclinicalBert (Out-of-domain)	✗	mean	**0.672**	**0.680**	0.543	0.632
MI-Zero (Ours)	HistPathGPT (None)	✓	mean	0.655	0.595	0.573	0.608
	HistPathGPT (In-domain)	✓	mean	0.625	0.590	**0.637**	0.617
	PubmedBert (Out-of-domain)	✓	mean	0.587	0.650	0.730	**0.656**
	BioclinicalBert (Out-of-domain)	✓	mean	**0.675**	**0.682**	0.543	0.634

In-domain text

Out-of-domain text

Slide-level zero-shot transfer: image pretraining

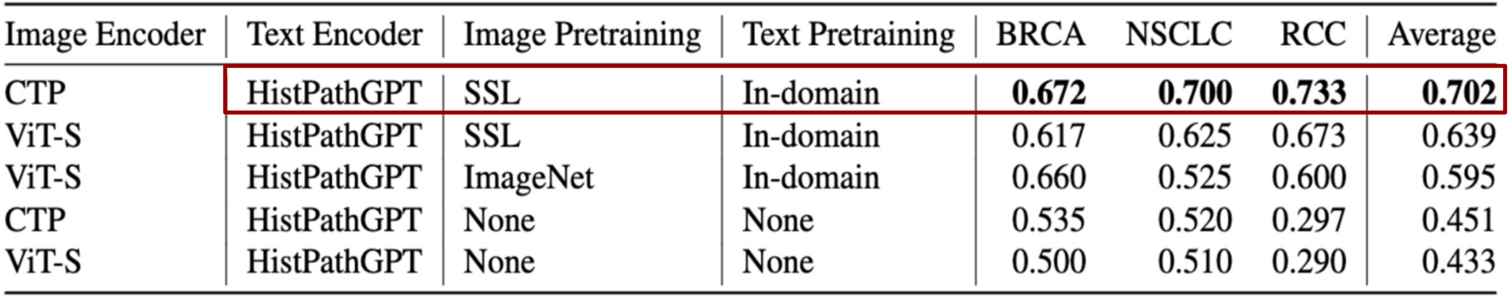
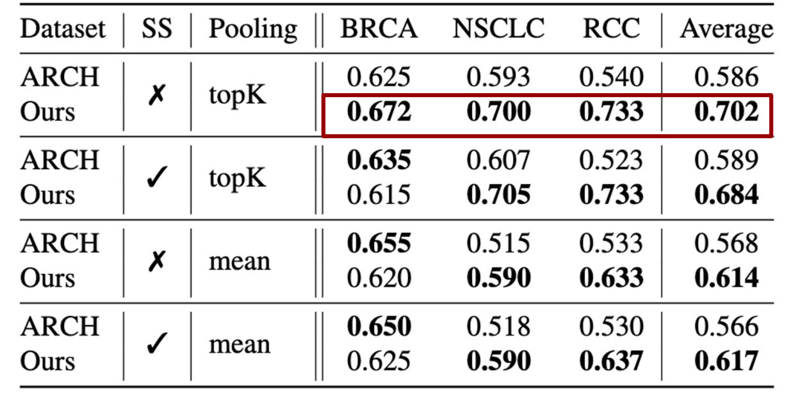


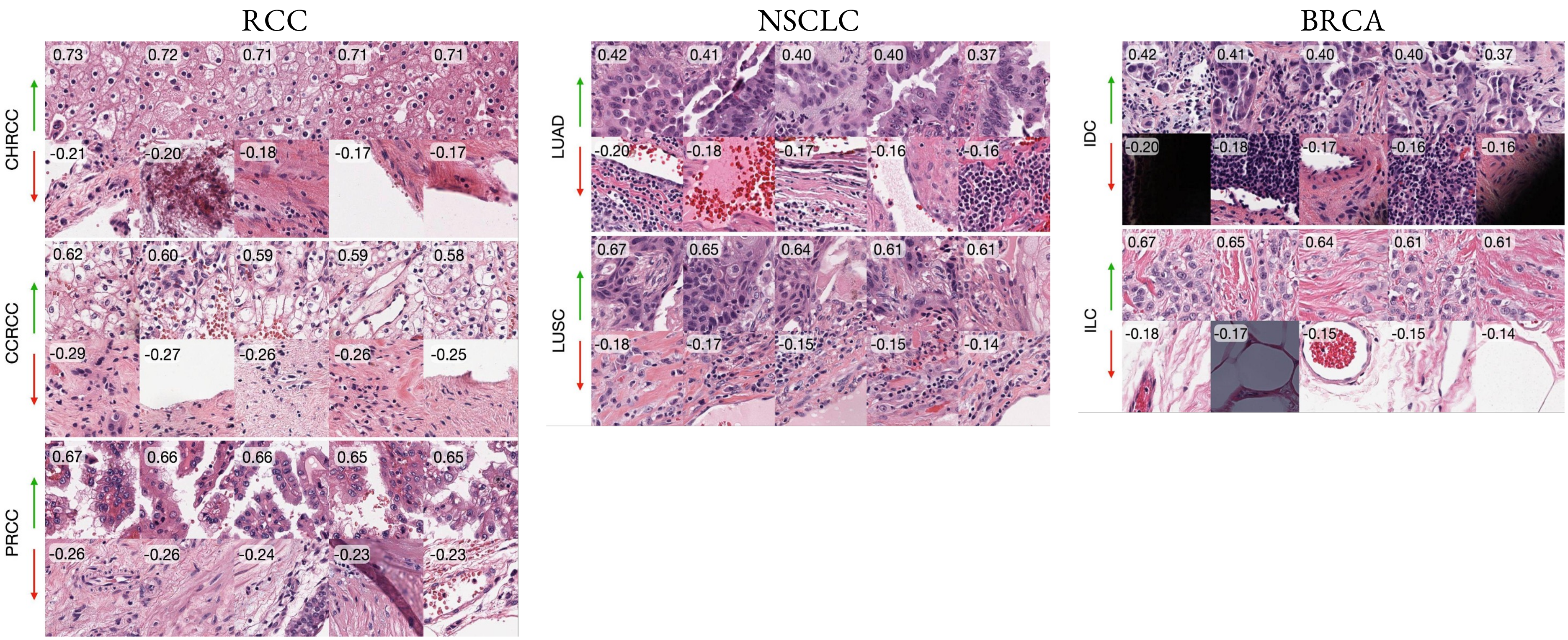
Image Encoder	Text Encoder	Image Pretraining	Text Pretraining	BRCA	NSCLC	RCC	Average
CTP	HistPathGPT	SSL	In-domain	**0.672**	**0.700**	**0.733**	**0.702**
ViT-S	HistPathGPT	SSL	In-domain	0.617	0.625	0.673	0.639
ViT-S	HistPathGPT	ImageNet	In-domain	0.660	0.525	0.600	0.595
CTP	HistPathGPT	None	None	0.535	0.520	0.297	0.451
ViT-S	HistPathGPT	None	None	0.500	0.510	0.290	0.433

Slide-level zero-shot transfer: dataset comparison



Dataset	SS	Pooling	BRCA	NSCLC	RCC	Average
ARCH	✗	topK	0.625	0.593	0.540	0.586
Ours			**0.672**	**0.700**	**0.733**	**0.702**
ARCH	✓	topK	**0.635**	0.607	0.523	0.589
Ours			0.615	**0.705**	**0.733**	**0.684**
ARCH	✗	mean	**0.655**	0.515	0.533	0.568
Ours			0.620	**0.590**	**0.633**	**0.614**
ARCH	✓	mean	**0.650**	0.518	0.530	0.566
Ours			0.625	**0.590**	**0.637**	**0.617**

Similarity scores select diagnostically relevant patches



Mahmood Lab
AI for Pathology

The Mahmood Lab

HARVARD MEDICAL SCHOOL

Mass General Brigham

BROAD INSTITUTE

Dana-Farber Cancer Institute

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA