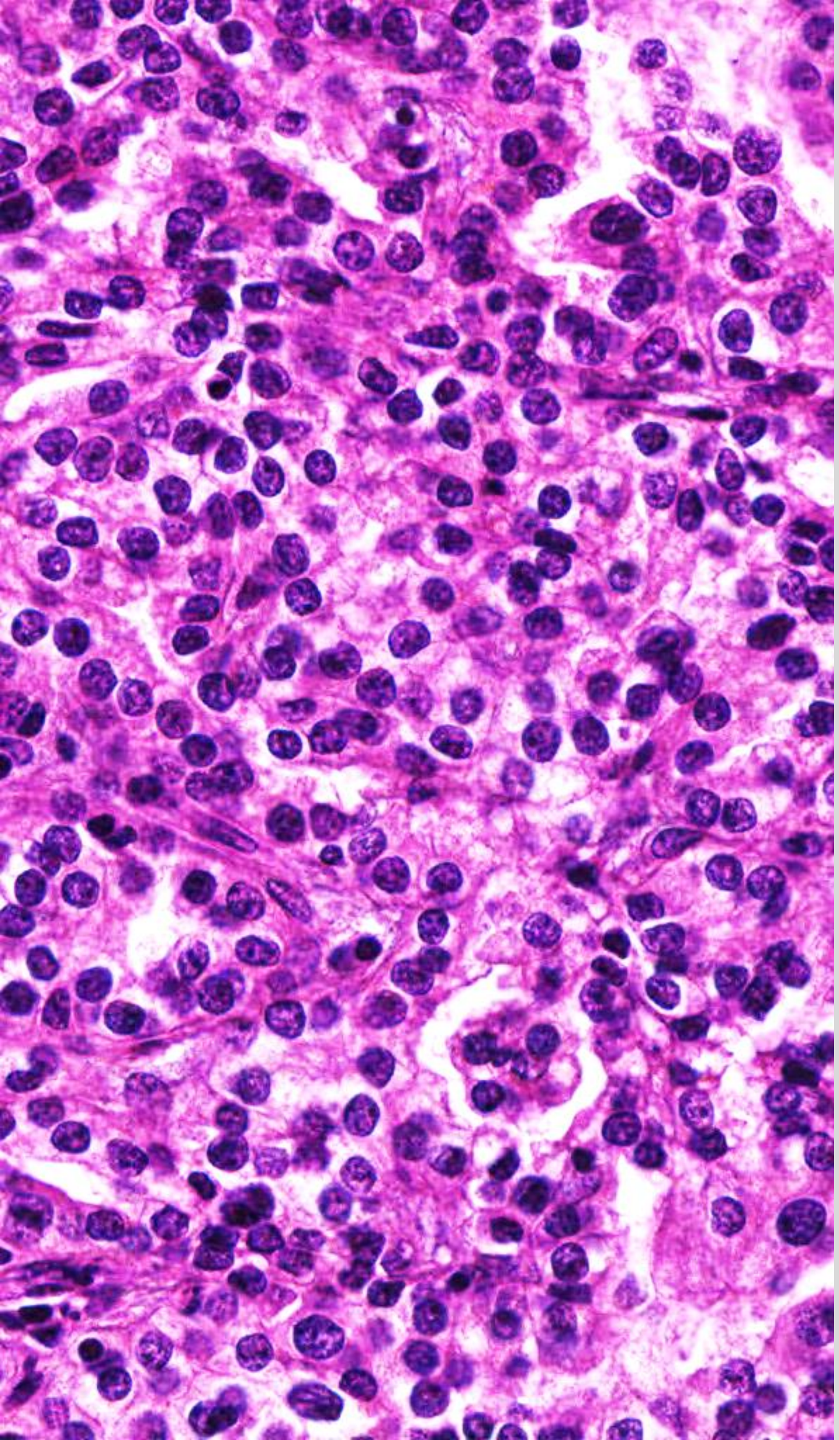
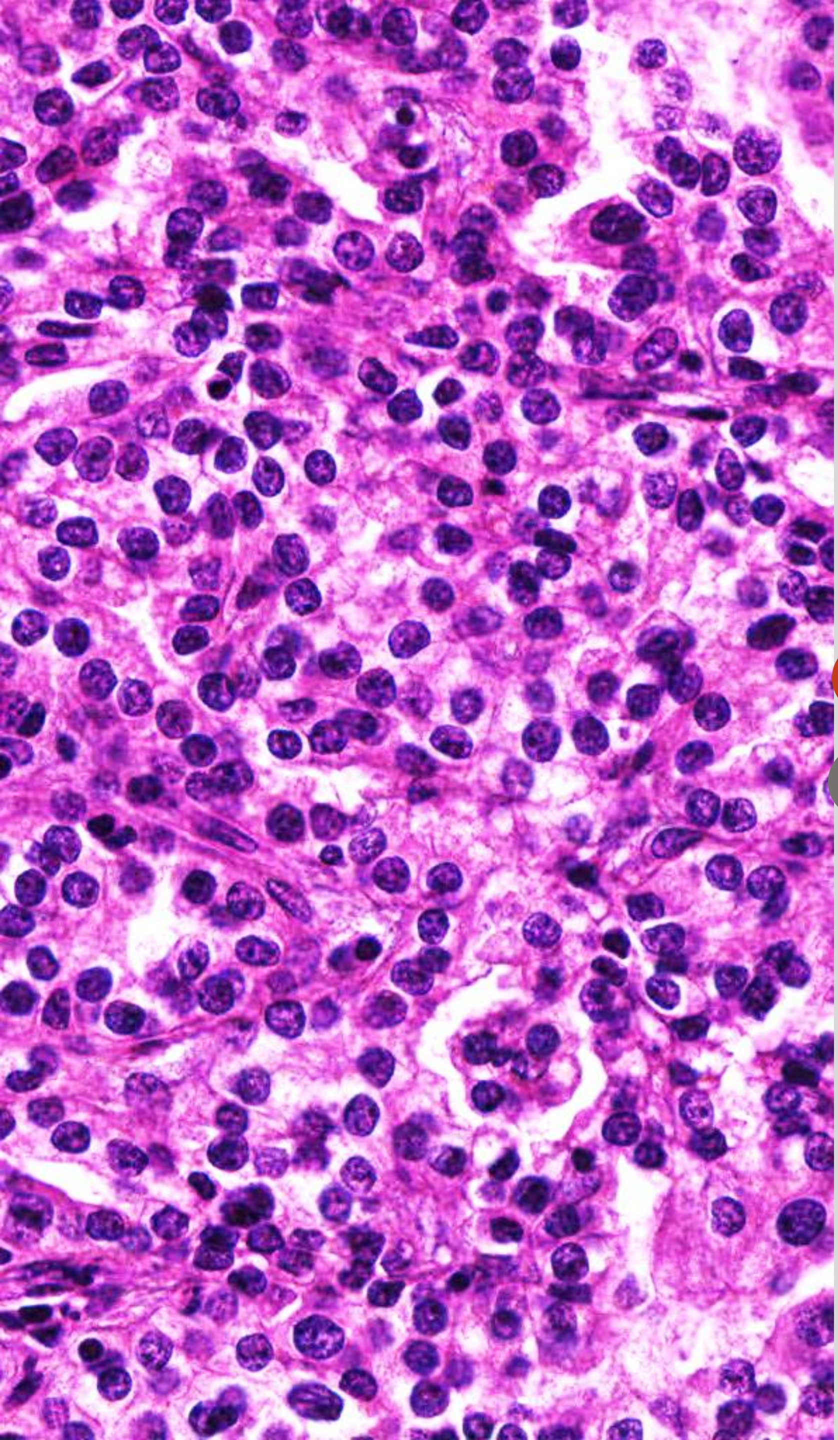


HEALNet:
Multimodal Fusion for Heterogeneous Biomedical Data

NeurIPS 2024

Konstantin Hemker, Nikola Simidjievski, Mateja Jamnik

UNIVERSITY OF CAMBRIDGE



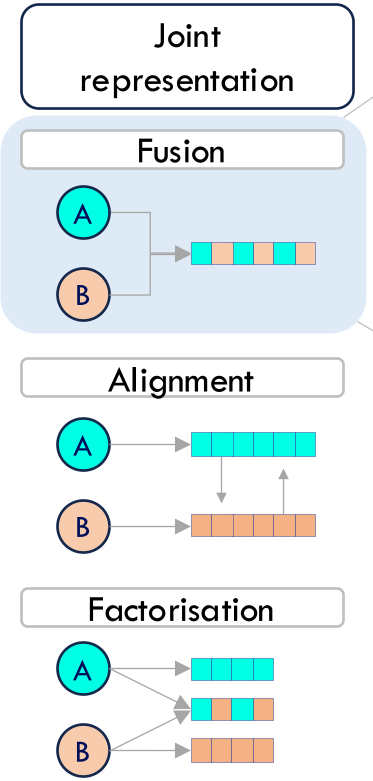
HEALNet: Multimodal Fusion for Heterogeneous Biomedical Data

NeurIPS 2024

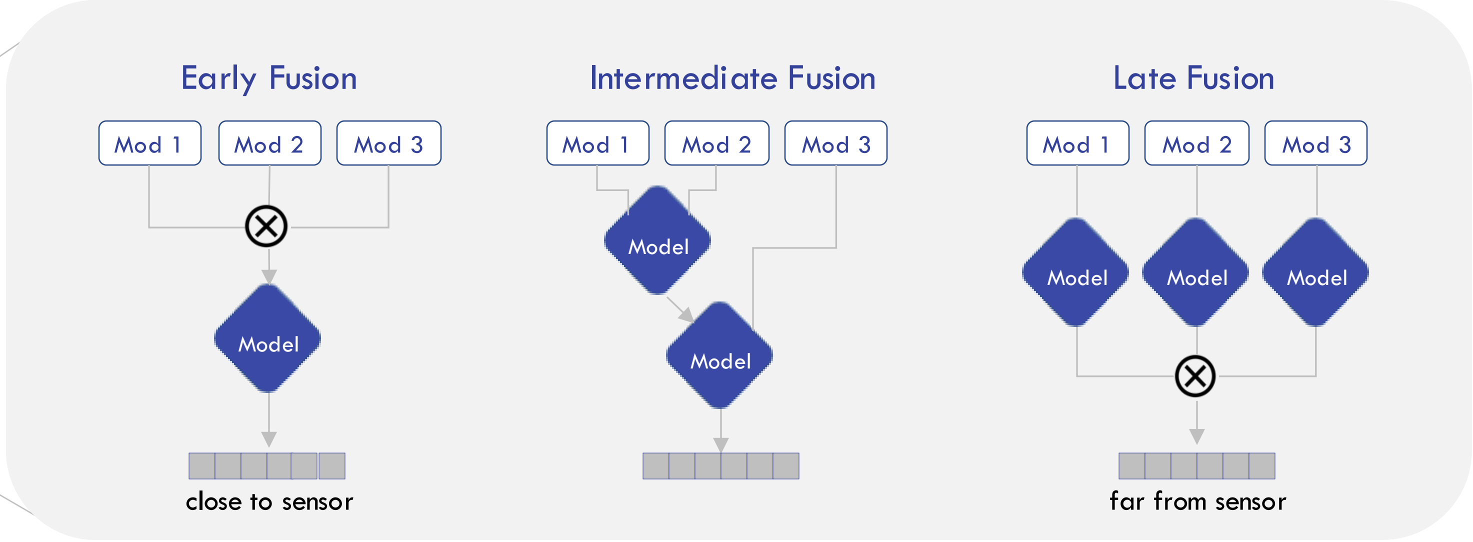
Konstantin Hemker, Nikola Simidjievski, Mateja Jamnik

UNIVERSITY OF CAMBRIDGE

Why fusion?



How do we encode uni- and cross-modal relationships?



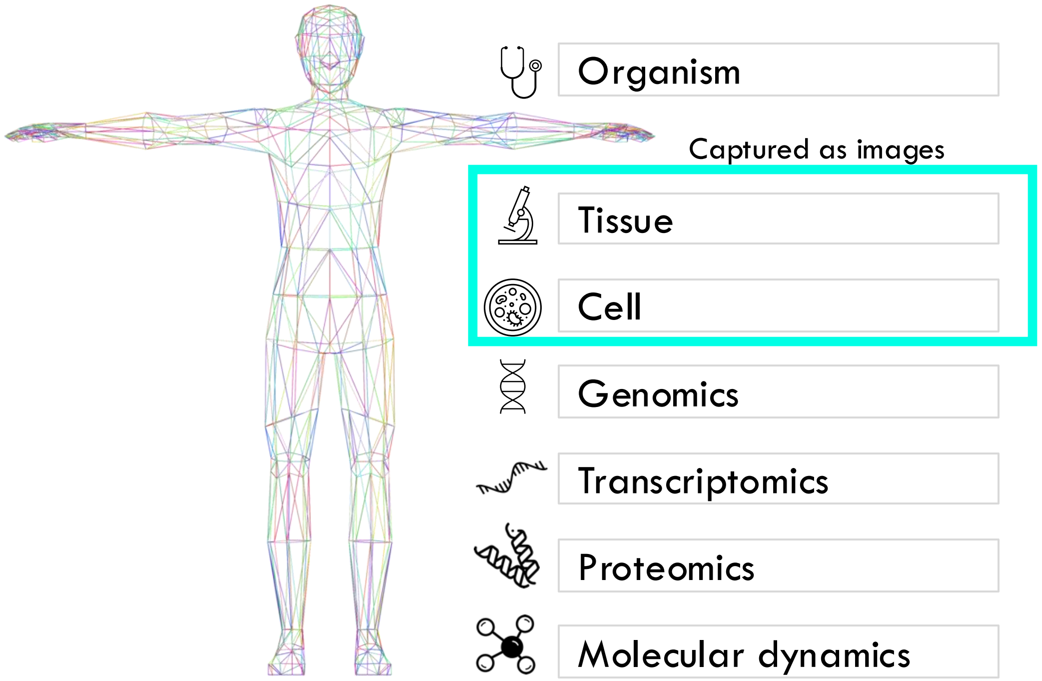
Common challenges

Often does not maintain modality structure	Most architectures are domain- and problem-specific	Does not capture cross-modal interactions

- Monolithic structure
- Requires paired data

Open challenges for fusion models in biomedical domains

Information from different biological scales can provide crucial predictive context…

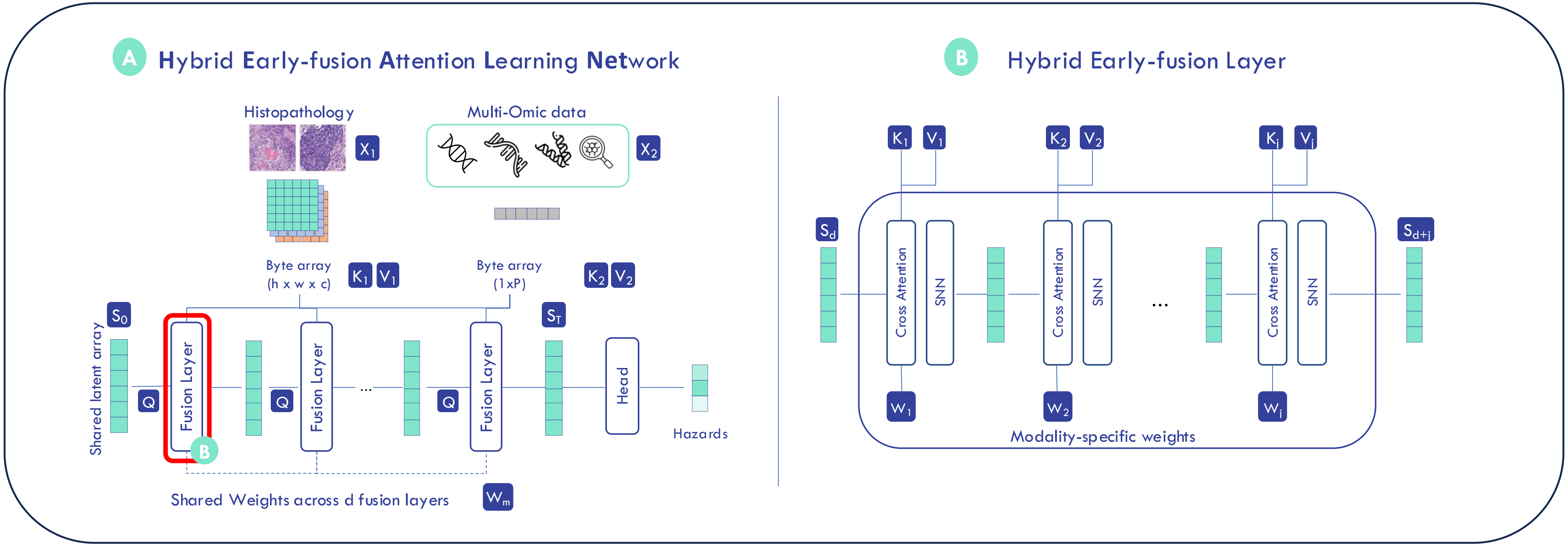


…but often require trading off modality-specific and shared information

Preserve structural signal	Find a joint representation that encodes the uni-modal **structural information** of each modality
Mutual context	Learn **cross-modal information** where one modality contextualises another modality, vice versa
Missing data	Handle **missing data** at training and inference time (often neglected, but highly relevant in clinical practice)
Lack of paired data	Often no clear modality **counterpart** between the different scales, one-to-many cardinalities
Discovery	**Explain** both uni-modal and cross-modal information that the model has learned

UNIVERSITY OF CAMBRIDGE

HEALNet uses shared and modality-specific parameter spaces



Preserve modality-specific structural information

Learn cross-modal interactions

Missing modalities & lack of paired data

Easy inspection using modality-specific attention weights

UNIVERSITY OF CAMBRIDGE

HEALNet preserves modality-specific structural information

Table 1: Mean and standard deviation of the concordance Index on four survival risk categories. We trained HEALNet and all baselines on four TCGA tasks and report the performance on the test set across five folds. HEALNet outperforms all of its multimodal baselines and three out of four unimodal baselines in absolute c-Index performance.

Model	BLCA	BRCA	KIRP	UCEC
Uni-modal (Omics)	0.606 ± 0.019	0.580 ± 0.027	0.780 ± 0.035	0.550 ± 0.026
Uni-modal (WSI)	0.556 ± 0.039	0.550 ± 0.037	0.533 ± 0.099	$\mathbf{0.630} \pm 0.028$
Porpoise (Late)	0.620 ± 0.048	0.630 ± 0.040	0.790 ± 0.041	0.590 ± 0.034
MCAT (Intermediate)	0.620 ± 0.040	0.589 ± 0.073	0.789 ± 0.087	0.589 ± 0.062
MOTCAT (Intermediate)	0.631 ± 0.051	0.607 ± 0.069	0.810 ± 0.062	0.587 ± 0.083
MultiModN (Sequential Fusion)	0.551 ± 0.060	0.582 ± 0.084	0.753 ± 0.152	0.610 ± 0.121
Perceiver (Early Fusion)	0.565 ± 0.042	0.566 ± 0.068	0.783 ± 0.135	0.623 ± 0.107
HEALNet (ours)	$\mathbf{0.668 \pm 0.036}$	$\mathbf{0.638 \pm 0.073}$	$\mathbf{0.812 \pm 0.055}$	$\mathit{0.626 \pm 0.037}$

Preserves modality-specific structural information

Learns cross-modal interactions

Effective handling of missing modalities

Easy inspection using modality-specific attention weights

UNIVERSITY OF CAMBRIDGE

Iterative attention setup allows to skip updates for missing data

Table 2: Analysis of the performance of HEALNet, trained on all modalities, in scenarios with missing modalities at inference, compared to unimodal baselines. Each test sample contains only one of either the Omic or WSI modality. The scenarios include test sets consisting of samples with only Omic modality, only WSI modality or a combination of both (at random). HEALNet achieves a higher c-Index across datasets, implying effective encoding of cross-modal information and handling different amounts of data with missing modalities.

Test	100% Omics		100% WSI		50%WSI + 50% Omics		WSI+Omic
	Uni-modal	HEALNet	Uni-modal	HEALNet	Uni-modal	HEALNet	HEALNet
BLCA	0.606	<u>0.618</u>	0.487	<u>0.501</u>	0.547	<u>0.612</u>	**0.668**
BRCA	0.556	<u>0.571</u>	0.529	<u>0.539</u>	0.543	<u>0.541</u>	**0.638**
KIRP	0.771	<u>0.773</u>	0.518	<u>0.526</u>	0.644	<u>0.714</u>	**0.812**
UCEC	0.509	<u>0.529</u>	0.558	<u>0.584</u>	0.533	<u>0.580</u>	**0.626**

Preserves modality-specific structural information

Learns cross-modal interactions

Effective handling of missing modalities

Easy inspection using modality-specific attention weights

Attention-based design allows *some* explainability

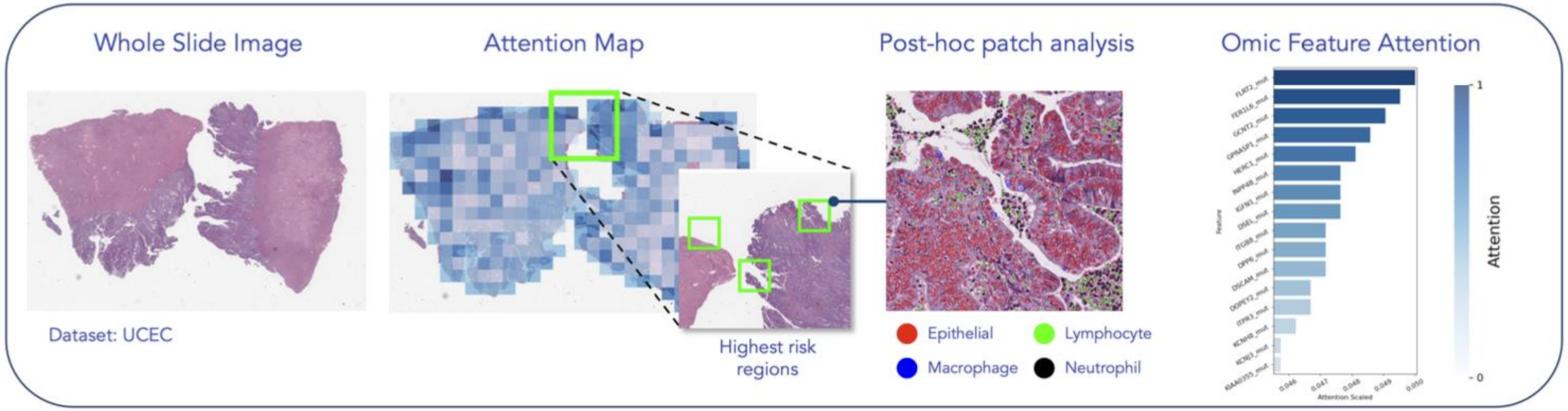


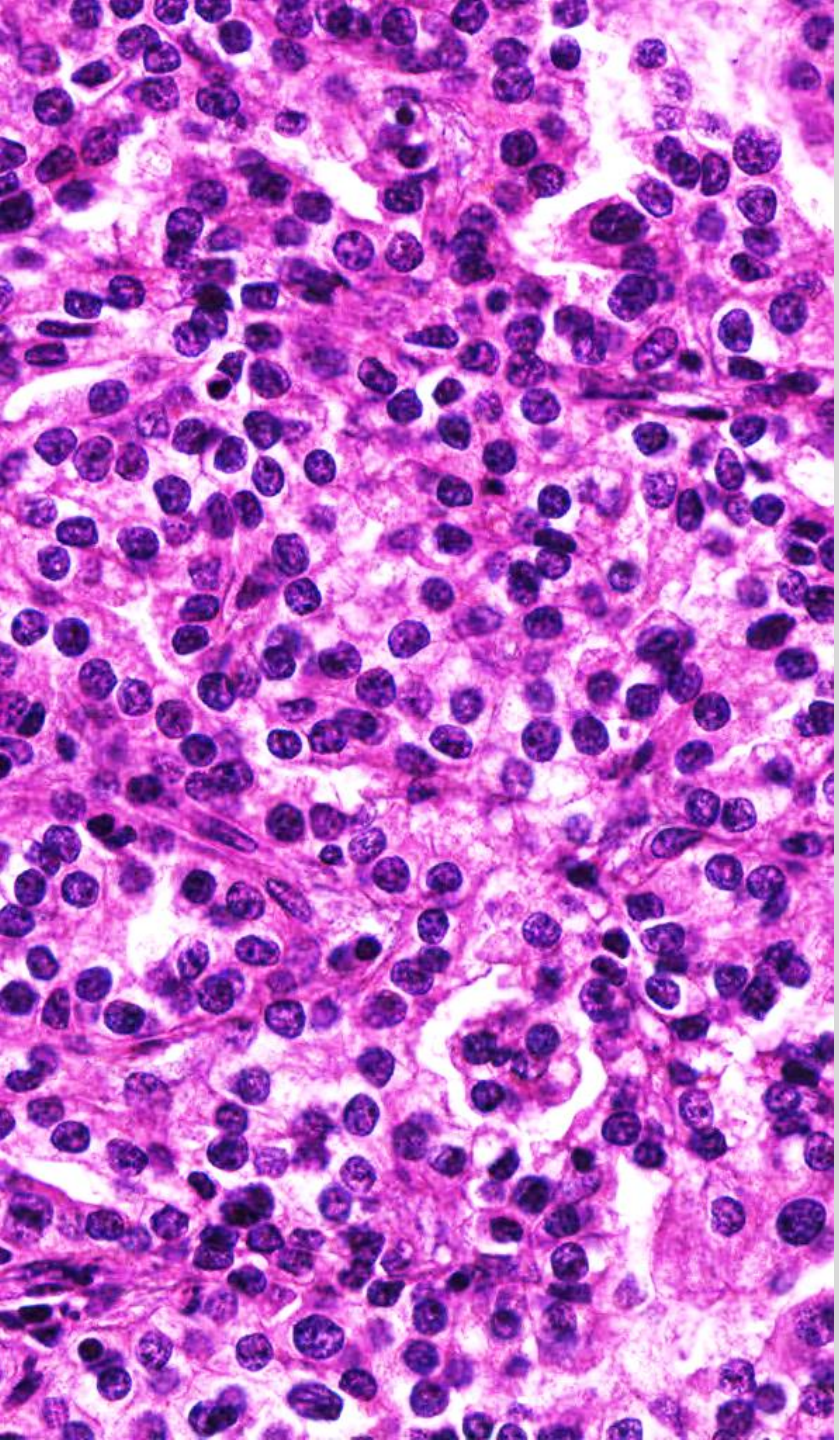
Figure 3: Illustration of model's inspection capabilities using HEALNet on a high-risk patient of the UCEC study. We use the mean modality-specific attention weights across layers to highlight high-risk regions and inspect high-attention omic features. Individual patches can be used for further clinical or computational post-hoc analysis such as nucleus segmentation. We observe that the high-risk regions exhibit a very high concentration and different arrangement of epithelial cells (red) which is commonly associated with the origin of various cancer types [Coradini et al., 2011].

Preserves modality-specific structural information

Learns cross-modal interactions

Effective handling of missing modalities

Easy inspection using modality-specific attention weights



Thank you

Code

Paper

Konstantin Hemker
PhD student in Multimodal Machine Learning
on Biomedical Data

Nikola Simidjievski
Senior Research Associate
Department of Oncology

Mateja Jamnik
Full Professor of Artificial Intelligence
Department of Computer Science &
Technology

konstantin.hemker@cl.cam.ac.uk

github.com/konsti-int-i/healnet

GATES Cambridge

UNIVERSITY OF CAMBRIDGE