

Synergistic Information Disentanglement for Omni-modal Slide Representation Learning in Computational Pathology

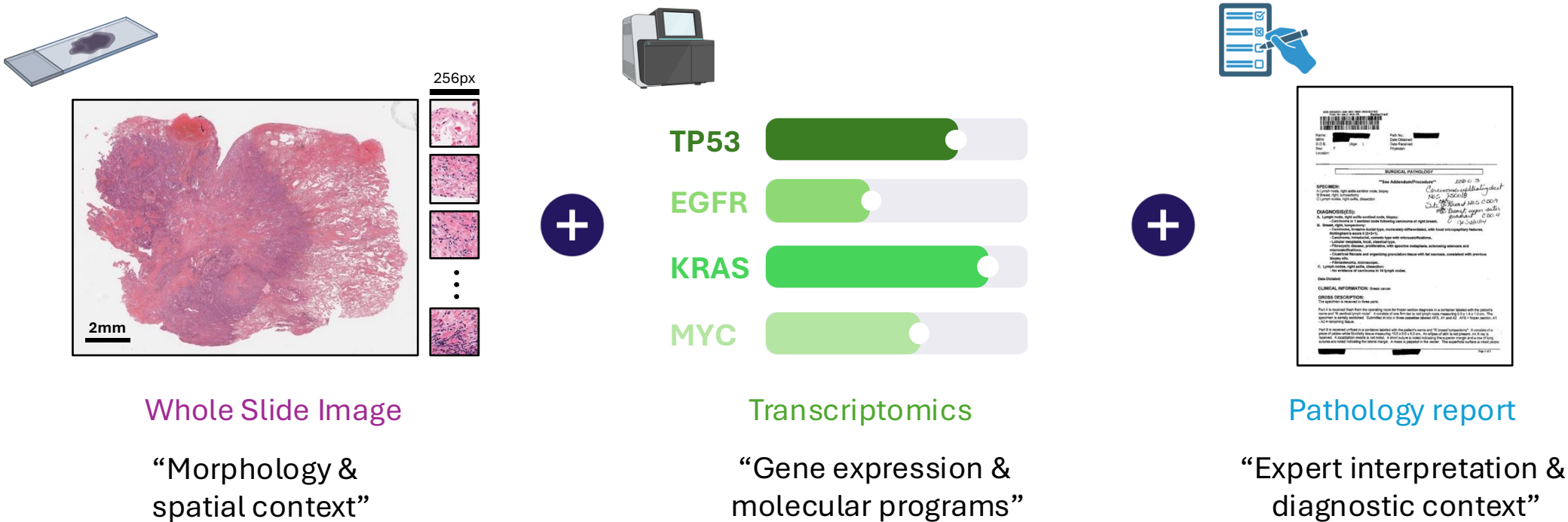
Mingxin Liu

Jiangsu Key Laboratory of Intelligent Medical Image Computing,
School of Artificial Intelligence,
Nanjing University of Information Science and Technology

MICCAI 2026 @ Strasbourg, France

Background

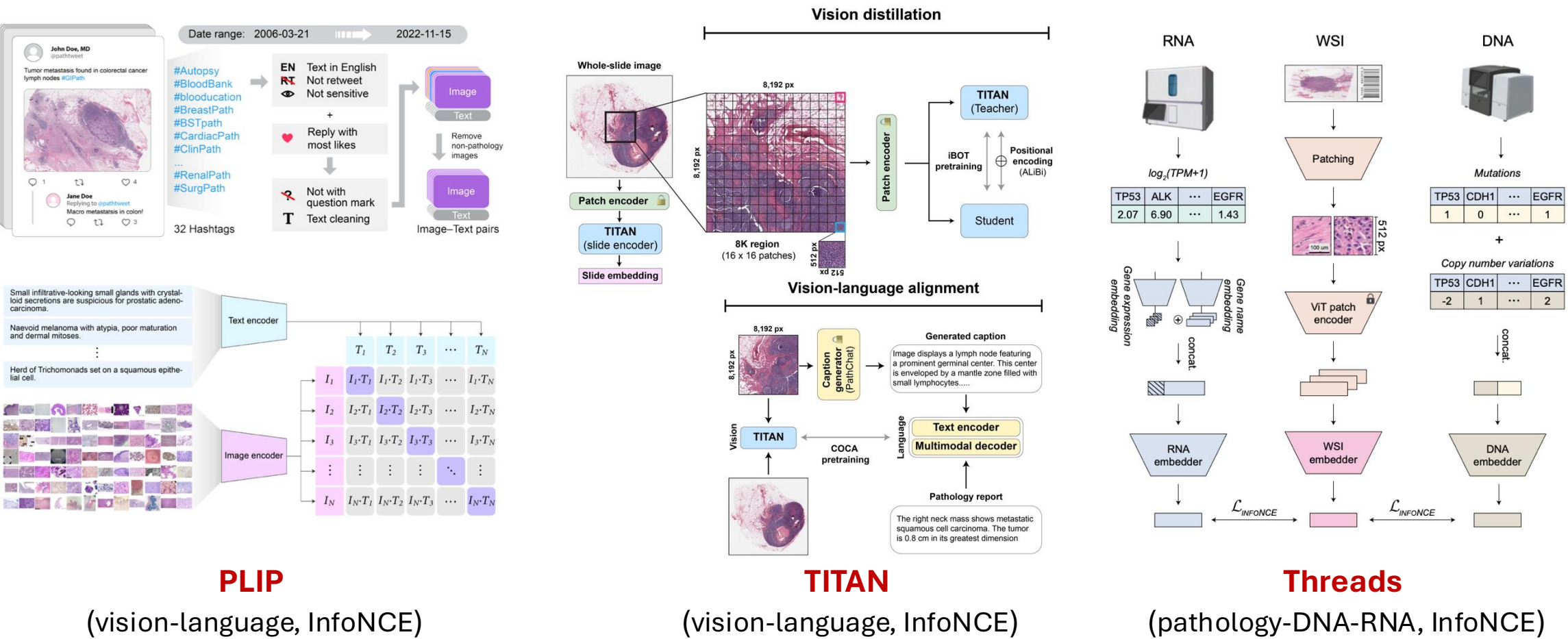
- Multiple modalities capture complementary evidence from the same patient



“Each modality reveals information that the others cannot fully provide”

What conventional multimodal learning optimizes?

- Many multimodal methods emphasize cross-modal contrastive alignment

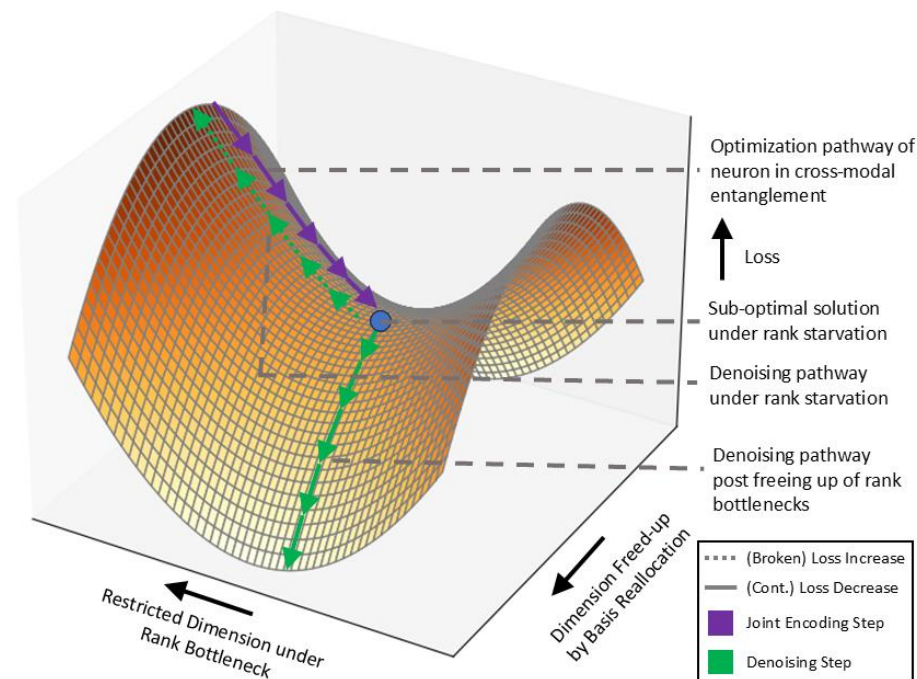
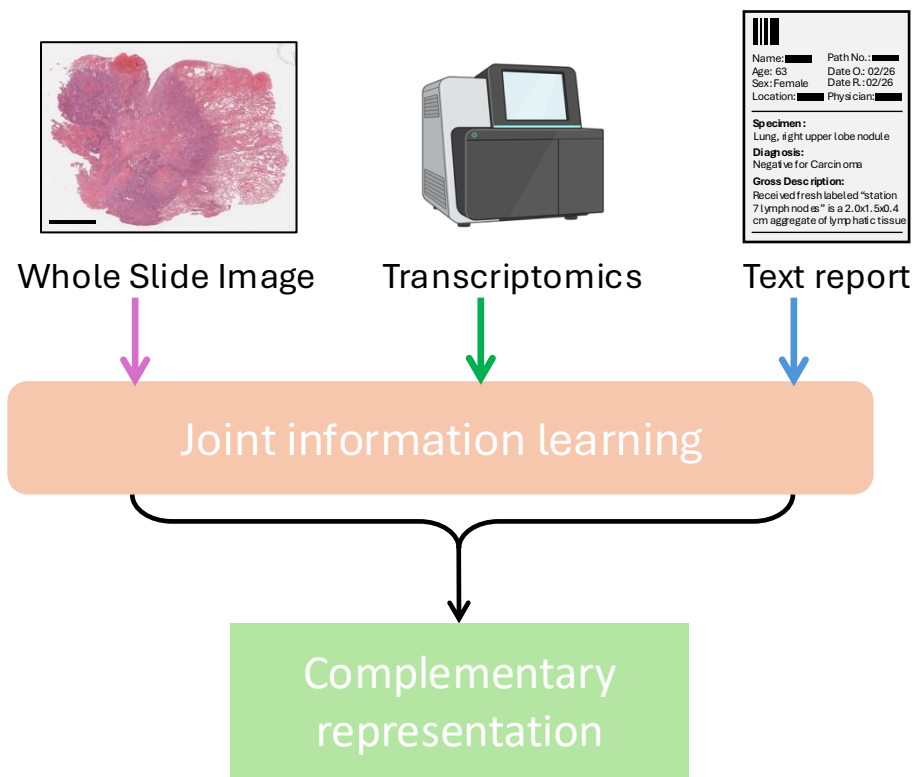


More modalities **do not** guarantee more useful information

- A model can be multimodal at the input, but rely on only a subset of modalities



“Ideally”

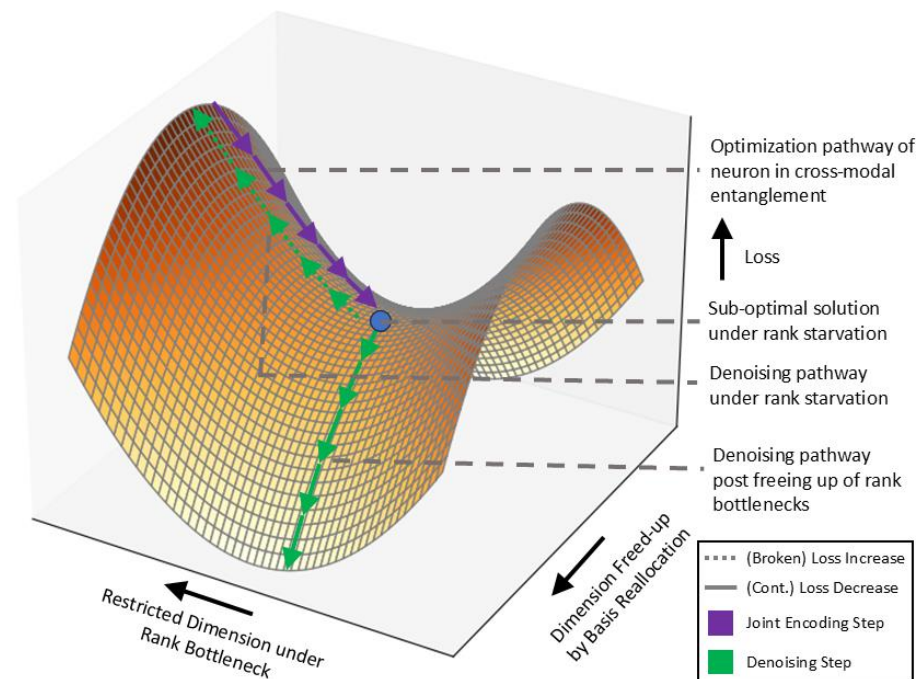
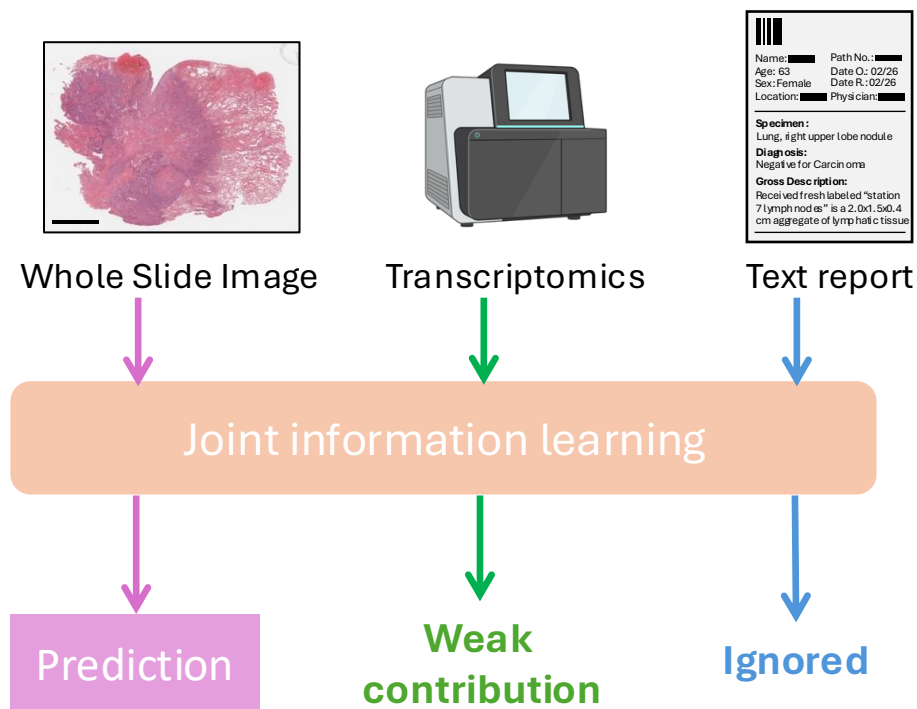


Modality collapse results from noisy features of one modality getting **entangled** with predictive features of another through polysemantic neurons in the fusion head.

More modalities **do not** guarantee more useful information

- A model can be multimodal at the input, but rely on only a subset of modalities

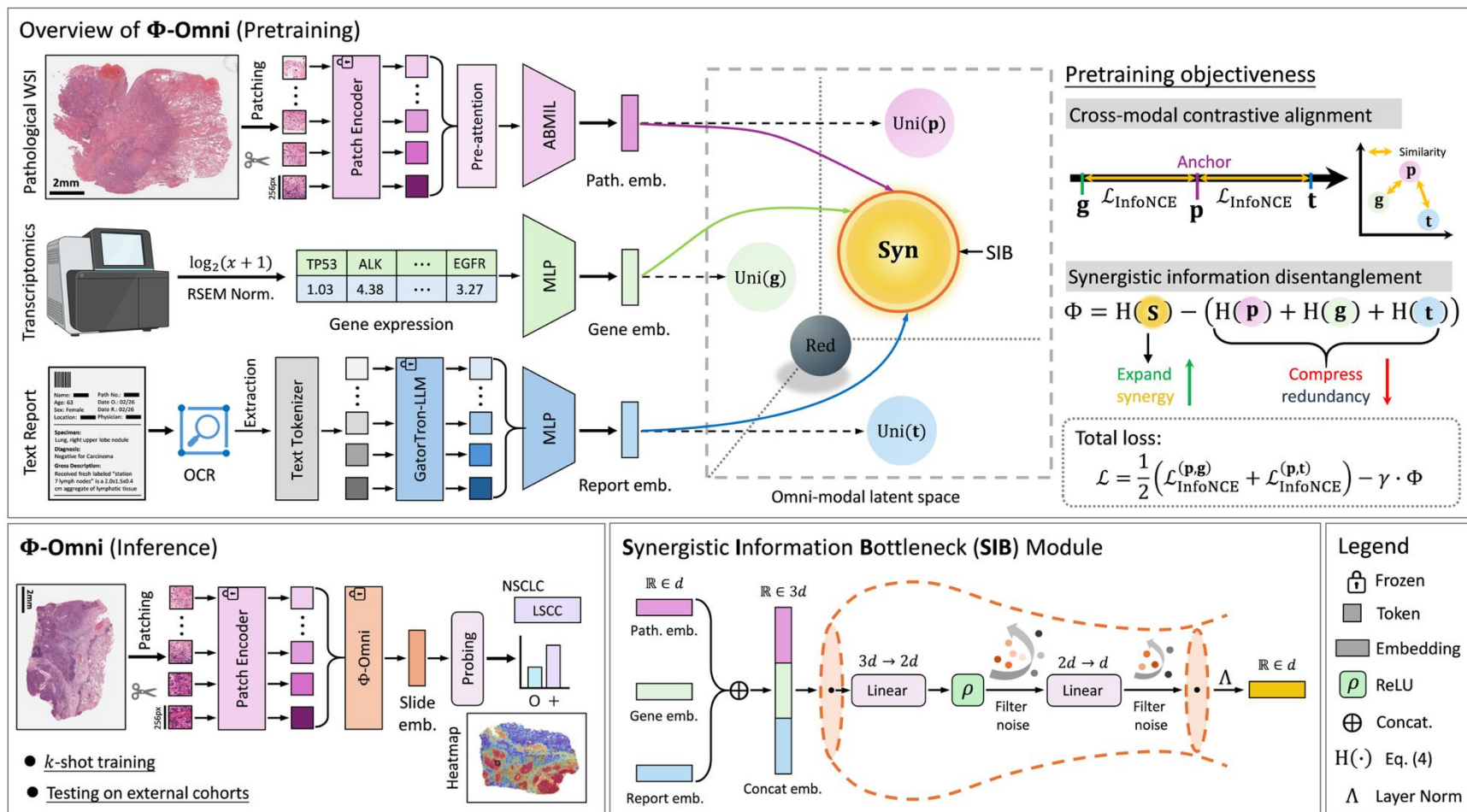
✗ “In practice, however...”



Modality collapse results from noisy features of one modality getting **entangled** with predictive features of another through polysemantic neurons in the fusion head.

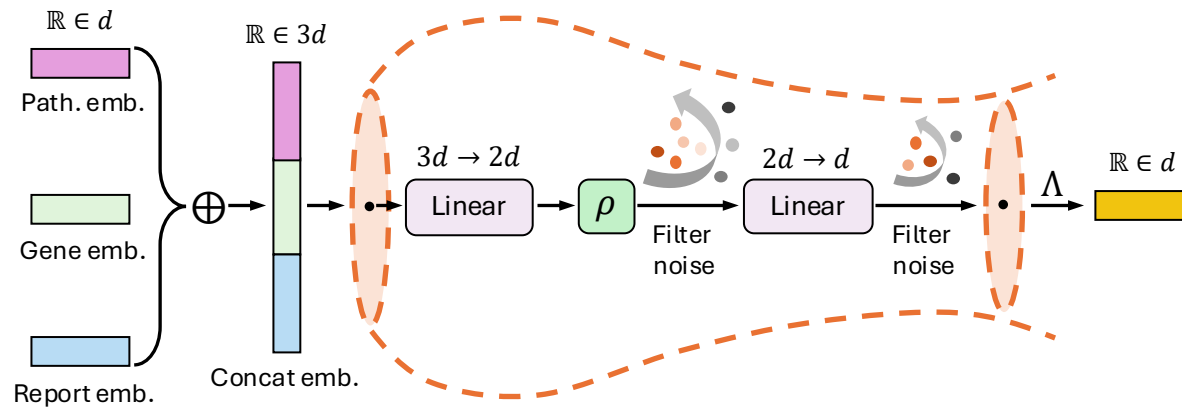
Our solution: Φ -Omni

- Learn slide representations from complementary multimodal information



Synergistic Information Bottleneck (SIB)

- SIB provides a compact space for multimodal interaction



Creating a dimensionality bottleneck: $3d \rightarrow 2d \rightarrow d$

$$\mathbf{Z}_{\text{mm}} = \mathcal{F}_{\text{SIB}}(\mathbf{X}_{\text{joint}})$$

$$\triangleq \text{LN}(\text{Proj}_2(\text{ReLU}(\text{Proj}_1[\mathbf{p}_i \oplus \mathbf{g}_i \oplus \mathbf{t}_i])))$$

- Concatenation gives access to all views
- Nonlinear projection enables interaction
- Dimensional reduction controls capacity
- Compression alone does not identify synergy

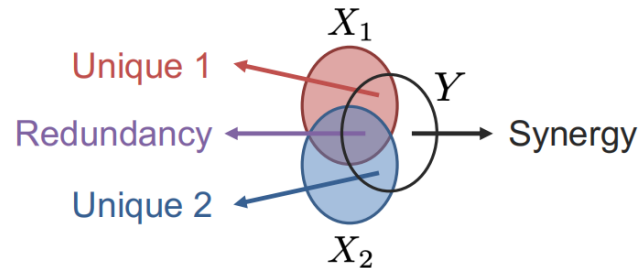
“Preserve useful information & Limit information capacity”

MI measures dependence; PID separates its sources

□ Mutual Information (MI)

□ Partial Information Decomposition (PID)

$$I(X_1, X_2; Z) = H(Z) - H(Z|X_1, X_2)$$



$$I(X_1, X_2; Z) = R + U_1 + U_2 + S$$

How much information?

Decompose

Where does the information come from?

R	U ₁	U ₂	S
Redundancy	Uniqueness	Uniqueness	Synergy
Shared cue	Only X ₁	Only X ₂	Joint only

“The equal MI can arise from fundamentally different multimodal contributions”

Joint information can exceed what individual views explain

- Pure synergy example

X_1	X_2	$Z = \text{XOR}$
0	0	0
0	1	1
1	0	1
1	1	0

* X_1 and X_2 are independent fair bits.

$$I(X_1; Z) = I(X_2; Z) = 0 \text{ bit}$$

$$I(X_1, X_2; Z) = 1 \text{ bit}$$

“XOR motivates a synergy–redundancy contrast”

- PID motivates a joint-minus-marginal contrast

* Two-source illustration.

$$I(X_1; Z) = R + U_1, I(X_2; Z) = R + U_2$$

$$I(X_1, X_2; Z) = R + U_1 + U_2 + S$$

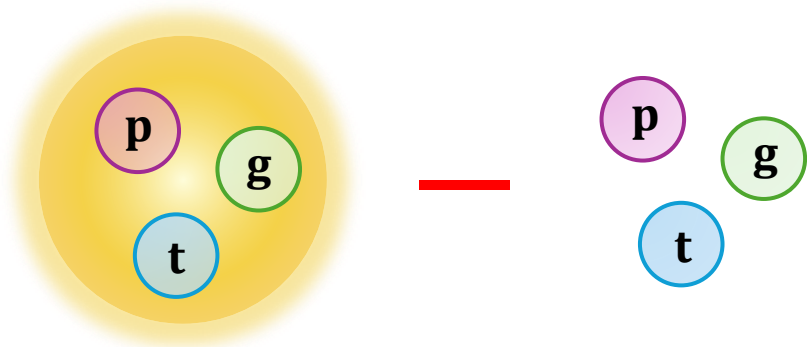
$$\begin{aligned}\Phi &= I(X_1, X_2; Z) - I(X_1; Z) - I(X_2; Z) \\ &= R + U_1 + U_2 + S - (R + U_1) - (R + U_2) \\ &= S - R\end{aligned}$$

“Joint information - sum of individual information”

“Positive Φ favors synergy relative to redundancy”

Synergistic Information Disentanglement

- Theory: whole-minus-sum



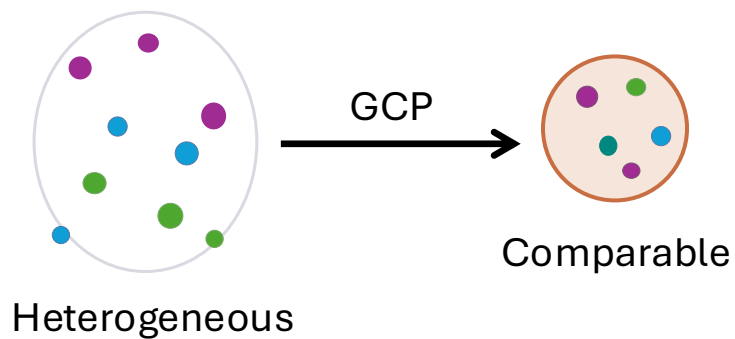
$$I(\mathbf{X}; \mathbf{Z}) = \text{Red}(\mathbf{X}; \mathbf{Z}) + \sum_{m \in \{p, g, t\}} \text{Uni}(\mathbf{X}_m; \mathbf{Z}) + \text{Syn}(\mathbf{X}; \mathbf{Z})$$

$$\Phi(\{p, g, t\} \rightarrow \mathbf{Z}_{\text{mm}})$$

$$\triangleq \underbrace{I(\mathbf{X}_{\text{joint}}; \mathbf{Z}_{\text{mm}})}_{\text{Total correlation}} - \underbrace{\sum_{m \in \{p, g, t\}} I(\mathbf{X}_m; \mathbf{Z}_{\text{mm}})}_{\text{Sum of marginal redundancy}}$$

- Gaussian Canonical Projector (GCP)

“makes volumes comparable”



$$\mathbf{Z}_* = \text{Sigmoid}(\text{LN}(\mathbf{W} \cdot \mathbf{h}_*))$$

$$H(\mathbf{Z}) = \frac{1}{2} \log \det \left(\frac{1}{B} \mathbf{Z}^T \mathbf{Z} + \epsilon \mathbf{I} \right)$$

* $\mathbf{I}$ is the identity matrix with $\epsilon = 10^{-5}$ for numerical stability.

Pre-training objective

□ Synergistic Information Disentanglement

$$\Phi = H(\mathbf{s}) - \underbrace{(H(\mathbf{p}) + H(\mathbf{g}) + H(\mathbf{t}))}_{\text{Compress redundancy}}$$

Expand synergy Compress redundancy

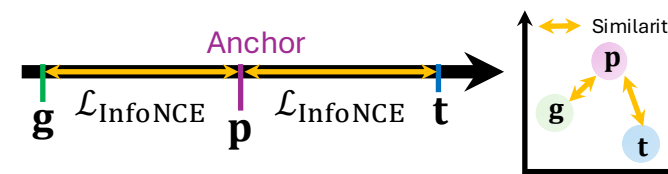
$$\mathcal{L}_{\Phi ID} = \sum_{m \in \{p, g, t\}} H(\mathbf{z}_m) - H(\mathbf{z}_{\text{joint}})$$

- Suppress unimodal redundancy
- Capture synergistic cross-modal interactions

□ Complete objectiveness:

$$\mathcal{L}_{\text{Total}} = \mathcal{L}_{\text{SymCL}} + \gamma \mathcal{L}_{\Phi ID}, \text{ where we set } \gamma = 0.2$$

□ Cross-modal contrastive alignment



$$\mathcal{L}_{\text{SymCL}} = \frac{1}{2} \sum_{\mathbf{k} \in \{\mathbf{g}, \mathbf{t}\}} (\ell(\mathbf{p}, \mathbf{k}) + \ell(\mathbf{k}, \mathbf{p}))$$

$$\ell(\mathbf{u}, \mathbf{v}) = -\frac{1}{B} \sum_{i=1}^B \frac{e^{\tau \mathbf{u}_i^T \mathbf{v}_i}}{\sum_{j=1}^M e^{\tau \mathbf{u}_j^T \mathbf{v}_j}}$$

Main results: multimodal training, unimodal deployment

- Multimodal supervision improves the deployed WSI representation

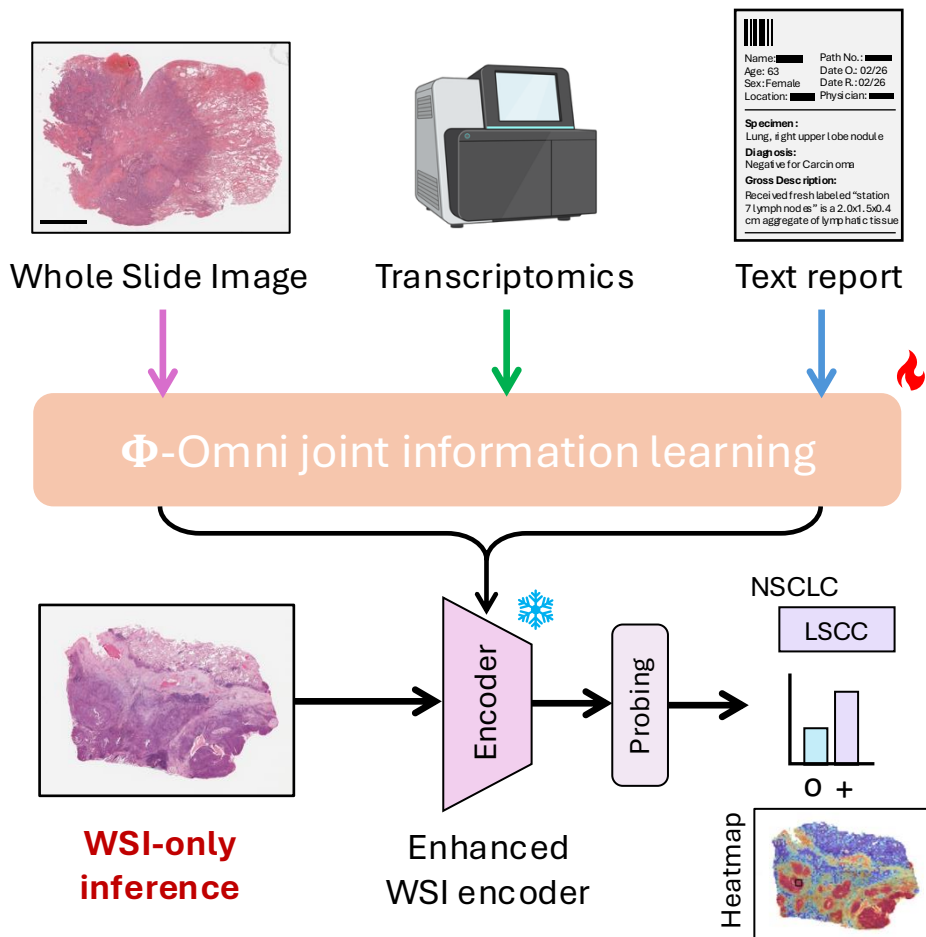


Table 1. Few-shot breast and lung cancer subtyping. Evaluation via macro-AUC on BRACS and CPTAC-NSCLC two cohorts. Best in **bold**, second best is underlined.

Model/Data		BRACS ($\uparrow$)				CPTAC-NSCLC ($\uparrow$)			
		$k=1$	$k=5$	$k=10$	$k=25$	$k=1$	$k=5$	$k=10$	$k=25$
MIL	ABMIL [5]	55.9 $\pm$ 2.8	64.8 $\pm$ 2.0	75.4 $\pm$ 2.6	82.3 $\pm$ 0.8	72.2 $\pm$ 12.4	91.9 $\pm$ 5.2	95.7 $\pm$ 2.7	98.3 $\pm$ 1.0
	CLAM [16]	55.5 $\pm$ 3.0	65.4 $\pm$ 3.2	74.9 $\pm$ 2.0	82.3 $\pm$ 1.2	74.3 $\pm$ 14.1	93.6 $\pm$ 3.1	96.5 $\pm$ 2.3	98.5 $\pm$ 0.7
	TransMIL [18]	57.1 $\pm$ 2.7	68.9 $\pm$ 2.2	75.6 $\pm$ 1.2	82.3 $\pm$ 1.9	71.7 $\pm$ 12.2	91.8 $\pm$ 3.0	95.2 $\pm$ 1.4	98.6 $\pm$ 0.5
	ILRA [26]	55.0 $\pm$ 1.8	62.5 $\pm$ 2.8	70.0 $\pm$ 1.6	79.0 $\pm$ 1.7	66.7 $\pm$ 15.9	87.1 $\pm$ 2.5	92.4 $\pm$ 1.7	97.0 $\pm$ 1.1
Linear probe	CTransPath* [24]	56.7 $\pm$ 2.8	63.3 $\pm$ 2.9	64.7 $\pm$ 1.4	68.9 $\pm$ 1.6	71.3 $\pm$ 11.1	88.0 $\pm$ 4.9	92.3 $\pm$ 2.1	96.0 $\pm$ 1.1
	UNiv2* [1]	55.8 $\pm$ 3.2	66.0 $\pm$ 2.0	70.6 $\pm$ 1.7	76.4 $\pm$ 1.5	72.9 $\pm$ 15.3	91.4 $\pm$ 4.2	95.3 $\pm$ 1.7	97.5 $\pm$ 1.0
	CONCH* [15]	58.4 $\pm$ 8.4	69.5 $\pm$ 2.8	73.4 $\pm$ 2.1	79.0 $\pm$ 1.7	83.1 $\pm$ 9.3	<u>95.8 $\pm$ 2.0</u>	<u>97.0 $\pm$ 0.7</u>	98.5 $\pm$ 0.5
	mSTAR* [27]	56.1 $\pm$ 3.0	66.0 $\pm$ 2.3	70.2 $\pm$ 2.0	75.4 $\pm$ 1.3	74.9 $\pm$ 11.9	92.1 $\pm$ 4.2	95.5 $\pm$ 1.7	97.8 $\pm$ 0.8
	GPFM* [17]	56.4 $\pm$ 2.7	66.2 $\pm$ 2.8	70.9 $\pm$ 2.0	75.8 $\pm$ 1.3	73.2 $\pm$ 10.8	90.3 $\pm$ 4.9	95.3 $\pm$ 1.9	97.8 $\pm$ 1.0
	CHIEF [25]	<u>64.2 $\pm$ 3.2</u>	73.6 $\pm$ 2.1	77.5 $\pm$ 1.4	81.9 $\pm$ 1.1	72.7 $\pm$ 10.9	91.0 $\pm$ 4.7	94.6 $\pm$ 1.6	97.4 $\pm$ 0.7
	TANGLE [6]	62.2 $\pm$ 2.3	<u>74.1 $\pm$ 2.4</u>	<u>78.1 $\pm$ 1.2</u>	<u>82.4 $\pm$ 0.8</u>	79.3 $\pm$ 10.2	95.3 $\pm$ 4.3	96.9 $\pm$ 2.1	<u>98.8 $\pm$ 0.2</u>
	MADELEINE [7]	61.3 $\pm$ 4.5	73.4 $\pm$ 2.4	76.7 $\pm$ 2.0	82.3 $\pm$ 1.3	81.3 $\pm$ 9.6	94.7 $\pm$ 2.7	96.6 $\pm$ 1.1	98.2 $\pm$ 0.6
	Φ -OMNI (Ours)	65.2 $\pm$ 2.6	76.5 $\pm$ 2.2	79.9 $\pm$ 1.0	83.3 $\pm$ 0.8	<u>81.7 $\pm$ 11.3</u>	97.5 $\pm$ 1.6	98.8 $\pm$ 0.5	99.4 $\pm$ 0.2

Table 2. Few-shot molecular status prediction. Evaluation via macro-AUC on three cohorts from CPTAC ($k=25$). Best in **bold**, second best is underlined.

	Model/Data	Reference Venue'Year	BRCA ($\uparrow$)		LUAD ($\uparrow$)		LSCC ($\uparrow$)		Avg. ($\uparrow$)
			PIK3CA	TP53	STK11	TP53	ARID1A	KEAP1	
MIL	ABMIL [5]	ICML'2018	61.2 $\pm$ 8.7	76.1 $\pm$ 4.6	90.8 $\pm$ 5.1	76.3 $\pm$ 5.0	82.2 $\pm$ 5.0	84.5 $\pm$ 3.8	78.5
	CLAM [16]	Nat. BME.'2021	60.5 $\pm$ 8.7	75.5 $\pm$ 5.2	92.1 $\pm$ 4.4	78.3 $\pm$ 3.9	81.0 $\pm$ 5.1	84.4 $\pm$ 5.0	78.6
	TransMIL [18]	NeurIPS'2021	59.4 $\pm$ 7.6	78.6 $\pm$ 3.5	91.3 $\pm$ 3.3	78.0 $\pm$ 4.8	80.9 $\pm$ 3.6	83.9 $\pm$ 5.2	78.7
	ILRA [26]	ICLR'2023	61.4 $\pm$ 7.9	74.5 $\pm$ 5.9	91.7 $\pm$ 4.3	77.8 $\pm$ 4.5	81.3 $\pm$ 4.0	87.1 $\pm$ 6.3	79.0
Linear probe	CTransPath* [24]	MedIA'2022	62.1 $\pm$ 4.8	73.1 $\pm$ 4.8	85.3 $\pm$ 3.4	70.8 $\pm$ 2.7	75.2 $\pm$ 7.5	82.4 $\pm$ 3.5	74.8
	UNiv2* [1]	Nat. Med.'2024	59.6 $\pm$ 5.2	77.2 $\pm$ 5.2	92.0 $\pm$ 2.4	77.7 $\pm$ 3.0	80.3 $\pm$ 7.3	87.6 $\pm$ 5.0	79.1
	CONCH* [15]	Nat. Med.'2024	56.2 $\pm$ 5.3	77.2 $\pm$ 5.0	89.2 $\pm$ 3.3	68.9 $\pm$ 4.1	73.2 $\pm$ 9.8	78.6 $\pm$ 5.2	73.9
	mSTAR* [27]	Nat. Comm.'2025	60.0 $\pm$ 4.8	78.1 $\pm$ 4.6	91.5 $\pm$ 2.6	77.2 $\pm$ 2.5	79.2 $\pm$ 7.0	87.5 $\pm$ 4.6	78.9
	GPFM* [17]	Nat. BME.'2025	59.9 $\pm$ 5.3	77.5 $\pm$ 4.2	91.9 $\pm$ 2.9	74.9 $\pm$ 3.0	78.8 $\pm$ 7.8	85.5 $\pm$ 4.5	78.1
	CHIEF [25]	Nature'2024	63.8 $\pm$ 6.2	78.1 $\pm$ 4.1	89.5 $\pm$ 3.2	74.0 $\pm$ 2.5	78.8 $\pm$ 7.3	81.7 $\pm$ 3.4	77.7
	TANGLE [6]	CVPR'2024	60.1 $\pm$ 6.2	81.0 $\pm$ 3.5	91.8 $\pm$ 2.7	79.7 $\pm$ 3.5	82.8 $\pm$ 5.3	85.6 $\pm$ 4.6	80.2
	MADELEINE [7]	ECCV'2024	61.4 $\pm$ 5.9	79.3 $\pm$ 4.6	92.1 $\pm$ 2.6	72.7 $\pm$ 2.5	75.9 $\pm$ 8.2	83.1 $\pm$ 6.4	77.4
	Φ -OMNI (Ours)	MICCAI'2026	64.0 $\pm$ 6.1	80.0 $\pm$ 3.6	93.3 $\pm$ 1.7	82.4 $\pm$ 2.5	85.1 $\pm$ 5.1	88.8 $\pm$ 5.2	82.3

Main results: multimodal training, unimodal deployment

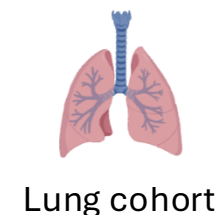
- Multimodal supervision improves the deployed WSI representation



Pretraining: TCGA-BRCA ($n=1,031$)

Evaluation:

- (1) BRACS ($n=547$, 7-way cancer subtyping)
- (2) CPTAC-BRCA ($n=112$, mutation prediction)



Pretraining: TCGA-NSCLC ($n=919$)

Evaluation:

- (1) CPTAC-NSCLC ($n=604$, cancer subtyping)
- (2) CPTAC-LUAD ($n=324$, mutation prediction)
- (3) CPTAC-LSCC ($n=304$, mutation prediction)

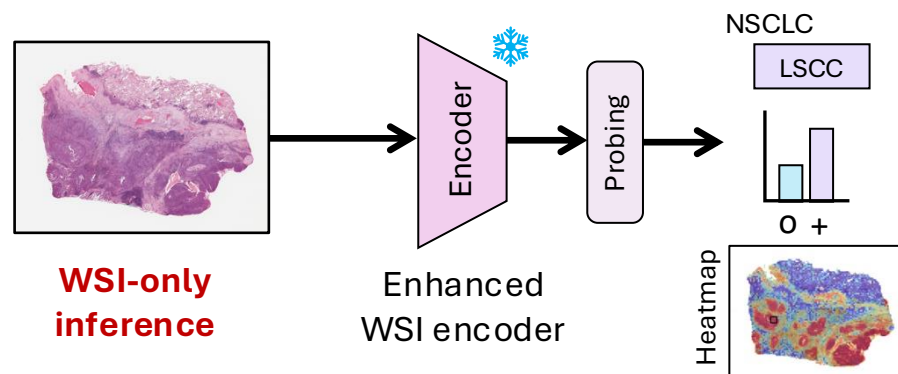


Table 1. Few-shot breast and lung cancer subtyping. Evaluation via macro-AUC on BRACS and CPTAC-NSCLC two cohorts. Best in **bold**, second best is underlined.

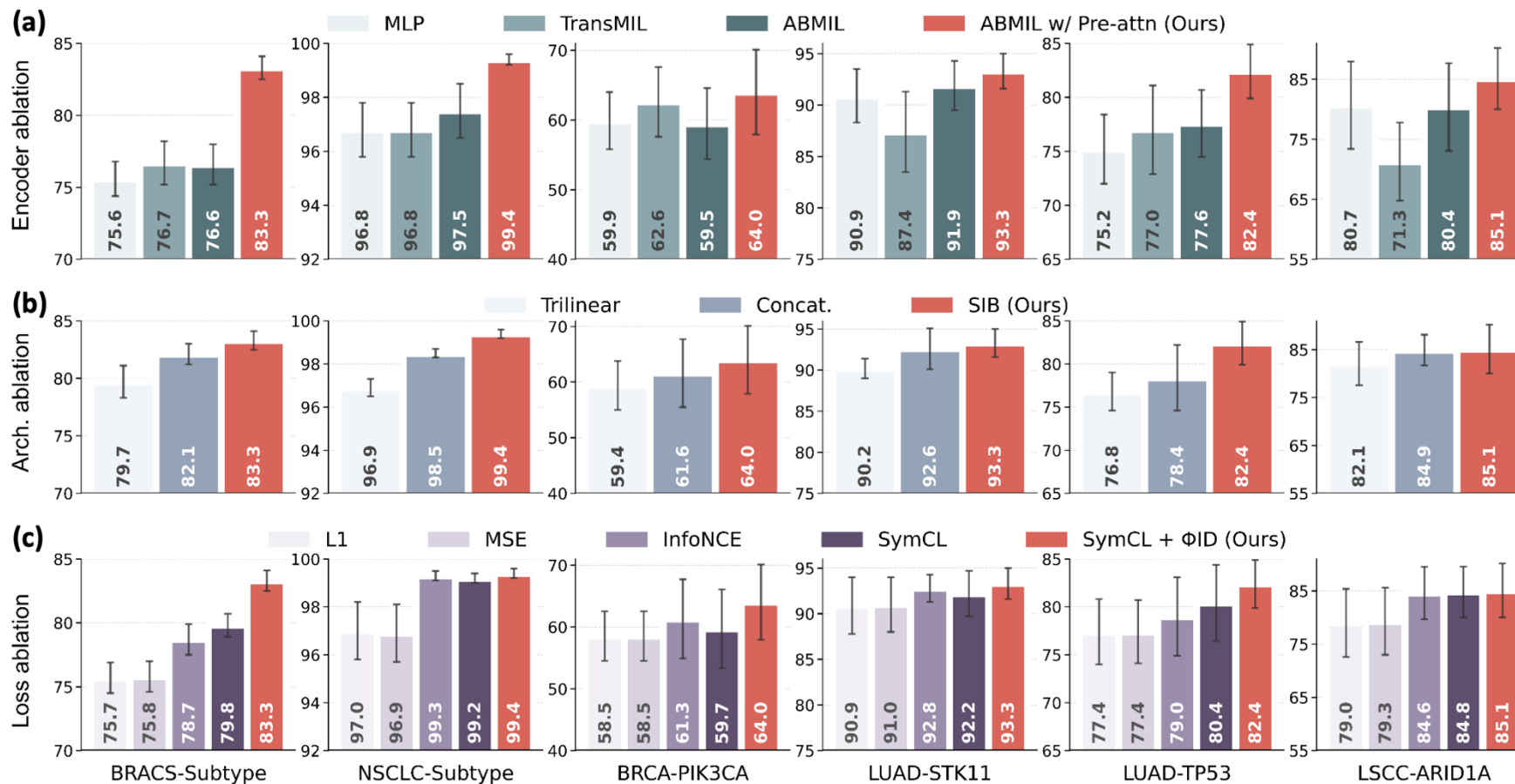
Model/Data		BRACS ($\uparrow$)				CPTAC-NSCLC ($\uparrow$)			
		$k=1$	$k=5$	$k=10$	$k=25$	$k=1$	$k=5$	$k=10$	$k=25$
MIL	ABMIL [5]	55.9 $\pm$ 2.8	64.8 $\pm$ 2.0	75.4 $\pm$ 2.6	82.3 $\pm$ 0.8	72.2 $\pm$ 12.4	91.9 $\pm$ 5.2	95.7 $\pm$ 2.7	98.3 $\pm$ 1.0
	CLAM [16]	55.5 $\pm$ 3.0	65.4 $\pm$ 3.2	74.9 $\pm$ 2.0	82.3 $\pm$ 1.2	74.3 $\pm$ 14.1	93.6 $\pm$ 3.1	96.5 $\pm$ 2.3	98.5 $\pm$ 0.7
	TransMIL [18]	57.1 $\pm$ 2.7	68.9 $\pm$ 2.2	75.6 $\pm$ 1.2	82.3 $\pm$ 1.9	71.7 $\pm$ 12.2	91.8 $\pm$ 3.0	95.2 $\pm$ 1.4	98.6 $\pm$ 0.5
	ILRA [26]	55.0 $\pm$ 1.8	62.5 $\pm$ 2.8	70.0 $\pm$ 1.6	79.0 $\pm$ 1.7	66.7 $\pm$ 15.9	87.1 $\pm$ 2.5	92.4 $\pm$ 1.7	97.0 $\pm$ 1.1
Linear probe	CTransPath* [24]	56.7 $\pm$ 2.8	63.3 $\pm$ 2.9	64.7 $\pm$ 1.4	68.9 $\pm$ 1.6	71.3 $\pm$ 11.1	88.0 $\pm$ 4.9	92.3 $\pm$ 2.1	96.0 $\pm$ 1.1
	UNiv2* [1]	55.8 $\pm$ 3.2	66.0 $\pm$ 2.0	70.6 $\pm$ 1.7	76.4 $\pm$ 1.5	72.9 $\pm$ 15.3	91.4 $\pm$ 4.2	95.3 $\pm$ 1.7	97.5 $\pm$ 1.0
	CONCH* [15]	58.4 $\pm$ 8.4	69.5 $\pm$ 2.8	73.4 $\pm$ 2.1	79.0 $\pm$ 1.7	83.1 $\pm$ 9.3	<u>95.8 $\pm$ 2.0</u>	<u>97.0 $\pm$ 0.7</u>	98.5 $\pm$ 0.5
	mSTAR* [27]	56.1 $\pm$ 3.0	66.0 $\pm$ 2.3	70.2 $\pm$ 2.0	75.4 $\pm$ 1.3	74.9 $\pm$ 11.9	92.1 $\pm$ 4.2	95.5 $\pm$ 1.7	97.8 $\pm$ 0.8
	GPfM* [17]	56.4 $\pm$ 2.7	66.2 $\pm$ 2.8	70.9 $\pm$ 2.0	75.8 $\pm$ 1.3	73.2 $\pm$ 10.8	90.3 $\pm$ 4.9	95.3 $\pm$ 1.9	97.8 $\pm$ 1.0
	CHIEF [25]	<u>64.2 $\pm$ 3.2</u>	73.6 $\pm$ 2.1	77.5 $\pm$ 1.4	81.9 $\pm$ 1.1	72.7 $\pm$ 10.9	91.0 $\pm$ 4.7	94.6 $\pm$ 1.6	97.4 $\pm$ 0.7
	TANGLE [6]	62.2 $\pm$ 2.3	74.1 $\pm$ 2.4	78.1 $\pm$ 1.2	82.4 $\pm$ 0.8	79.3 $\pm$ 10.2	95.3 $\pm$ 4.3	96.9 $\pm$ 2.1	<u>98.8 $\pm$ 0.2</u>
	MADELEINE [7]	61.3 $\pm$ 4.5	73.4 $\pm$ 2.4	76.7 $\pm$ 2.0	82.3 $\pm$ 1.3	81.3 $\pm$ 9.6	94.7 $\pm$ 2.7	96.6 $\pm$ 1.1	98.2 $\pm$ 0.6
	Φ -OMNI (Ours)	65.2 $\pm$ 2.6	76.5 $\pm$ 2.2	79.9 $\pm$ 1.0	83.3 $\pm$ 0.8	81.7 $\pm$ 11.3	97.5 $\pm$ 1.6	98.8 $\pm$ 0.5	99.4 $\pm$ 0.2

Table 2. Few-shot molecular status prediction. Evaluation via macro-AUC on three cohorts from CPTAC ($k=25$). Best in **bold**, second best is underlined.

	Model/Data	Reference Venue/Year	BRCA ($\uparrow$)		LUAD ($\uparrow$)		LSCC ($\uparrow$)		Avg. ($\uparrow$)
			PIK3CA	TP53	STK11	TP53	ARID1A	KEAP1	
MIL	ABMIL [5]	ICML'2018	61.2 $\pm$ 8.7	76.1 $\pm$ 4.6	90.8 $\pm$ 5.1	76.3 $\pm$ 5.0	82.2 $\pm$ 5.0	84.5 $\pm$ 3.8	78.5
	CLAM [16]	Nat. BME.'2021	60.5 $\pm$ 8.7	75.5 $\pm$ 5.2	<u>92.1 $\pm$ 4.4</u>	78.3 $\pm$ 3.9	81.0 $\pm$ 5.1	84.4 $\pm$ 5.0	78.6
	TransMIL [18]	NeurIPS'2021	59.4 $\pm$ 7.6	78.6 $\pm$ 3.5	91.3 $\pm$ 3.3	78.0 $\pm$ 4.8	80.9 $\pm$ 3.6	83.9 $\pm$ 5.2	78.7
	ILRA [26]	ICLR'2023	61.4 $\pm$ 7.9	74.5 $\pm$ 5.9	91.7 $\pm$ 4.3	77.8 $\pm$ 4.5	81.3 $\pm$ 4.0	87.1 $\pm$ 6.3	79.0
Linear probe	CTransPath* [24]	MedIA'2022	62.1 $\pm$ 4.8	73.1 $\pm$ 4.8	85.3 $\pm$ 3.4	70.8 $\pm$ 2.7	75.2 $\pm$ 7.5	82.4 $\pm$ 3.5	74.8
	UNiv2* [1]	Nat. Med.'2024	59.6 $\pm$ 5.2	77.2 $\pm$ 5.2	92.0 $\pm$ 2.4	77.7 $\pm$ 3.0	80.3 $\pm$ 7.3	<u>87.6 $\pm$ 5.0</u>	79.1
	CONCH* [15]	Nat. Med.'2024	56.2 $\pm$ 5.3	77.2 $\pm$ 5.0	89.2 $\pm$ 3.3	68.9 $\pm$ 4.1	73.2 $\pm$ 9.8	78.6 $\pm$ 5.2	73.9
	mSTAR* [27]	Nat. Comm.'2025	60.0 $\pm$ 4.8	78.1 $\pm$ 4.6	91.5 $\pm$ 2.6	77.2 $\pm$ 2.5	79.2 $\pm$ 7.0	87.5 $\pm$ 4.6	78.9
	GPfM* [17]	Nat. BME.'2025	59.9 $\pm$ 5.3	77.5 $\pm$ 4.2	91.9 $\pm$ 2.9	74.9 $\pm$ 3.0	78.8 $\pm$ 7.8	85.5 $\pm$ 4.5	78.1
	CHIEF [25]	Nature'2024	<u>63.8 $\pm$ 6.2</u>	78.1 $\pm$ 4.1	89.5 $\pm$ 3.2	74.0 $\pm$ 2.5	78.8 $\pm$ 7.3	81.7 $\pm$ 3.4	77.7
	TANGLE [6]	CVPR'2024	60.1 $\pm$ 6.2	81.0 $\pm$ 3.5	91.8 $\pm$ 2.7	<u>79.7 $\pm$ 3.5</u>	<u>82.8 $\pm$ 5.3</u>	85.6 $\pm$ 4.6	<u>80.2</u>
	MADELEINE [7]	ECCV'2024	61.4 $\pm$ 5.9	79.3 $\pm$ 4.6	<u>92.1 $\pm$ 2.6</u>	72.7 $\pm$ 2.5	75.9 $\pm$ 8.2	83.1 $\pm$ 6.4	77.4
	Φ -OMNI (Ours)	MICCAI'2026	64.0 $\pm$ 6.1	80.0 $\pm$ 3.6	93.3 $\pm$ 1.7	82.4 $\pm$ 2.5	85.1 $\pm$ 5.1	88.8 $\pm$ 5.2	82.3

Ablation study

□ Each component addresses a distinct failure mode



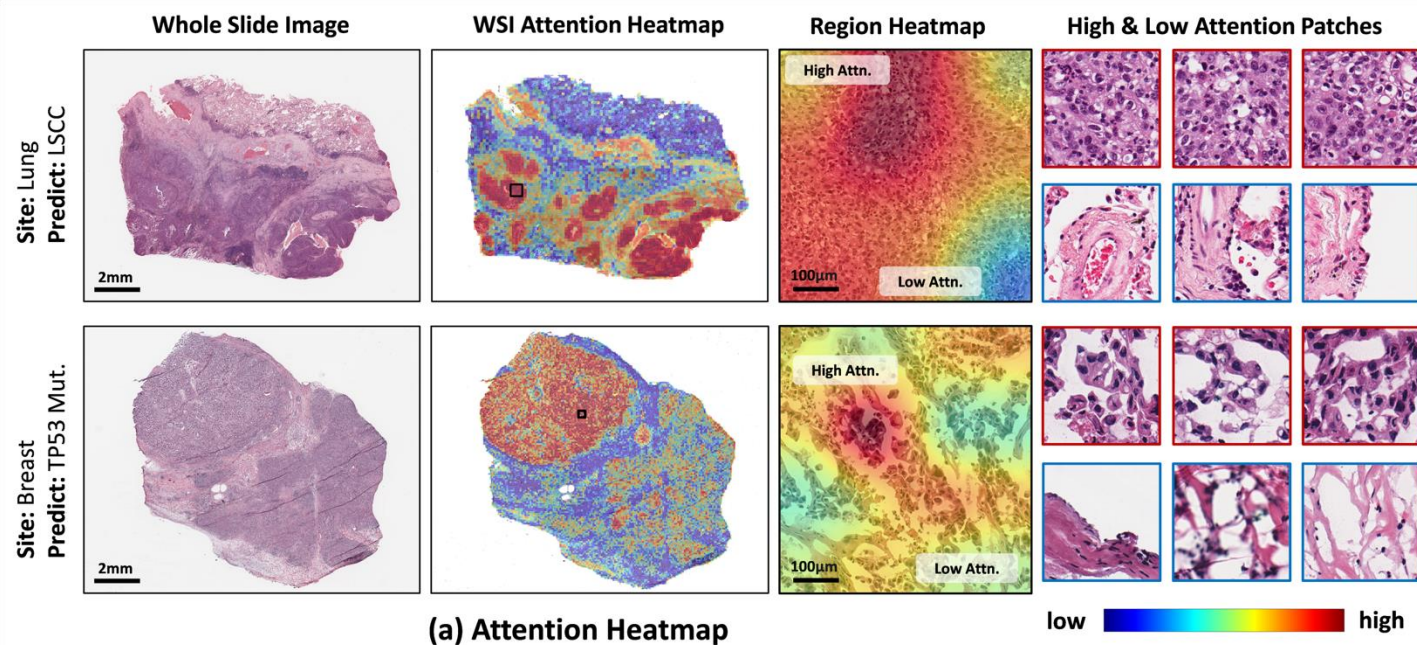
“ABMIL w/ pre-attention provides the most consistent slide encoder”

“SIB constrains the capacity of the multimodal transition”

“ Φ ID objective preserves the complementary information”

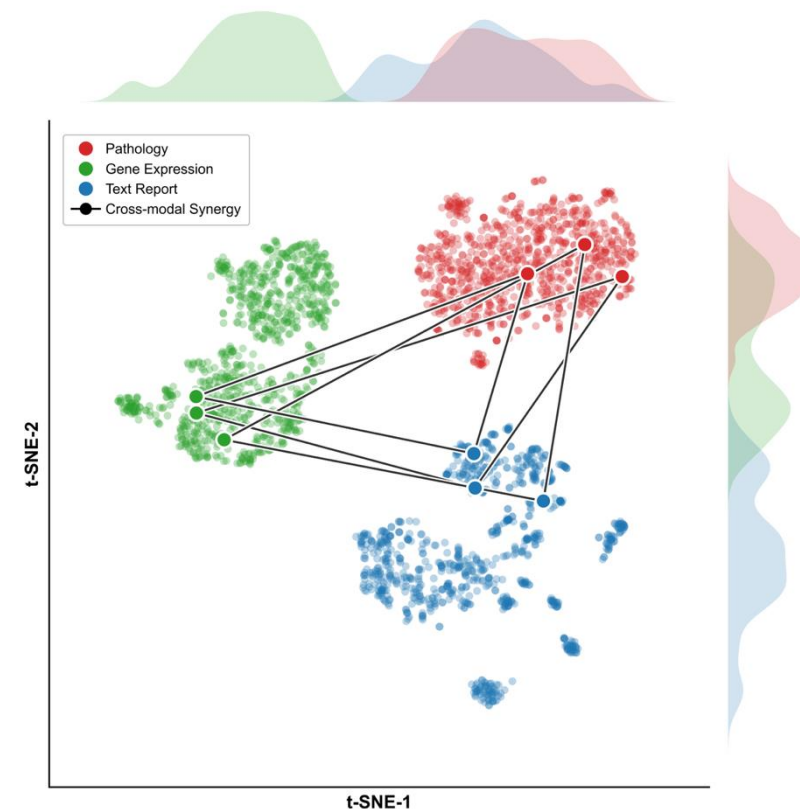
Interpretability Analysis

☐ Attention heatmaps



“Attention highlights morphologically relevant regions”

☐ Omni-modal disentanglement



“Preserves strong semantic alignment across heterogeneous modalities”

Take-home message

- ☐ Multimodal learning should preserve what emerges from interaction
 - ☐ More modalities are not enough
 - ☐ Information type matters
 - ☐ Φ -Omni makes complementarity transferable

A scenic view of a European city at sunset. In the foreground, a river flows through the city. On the left bank, there are several stone towers and buildings. A bridge crosses the river in the middle ground. In the background, a large cathedral with a tall spire is visible. The sky is filled with colorful clouds from the setting sun.

Thank you!

Mingxin Liu

Ph.D. student, IMIC Lab @ NUIST
Email: mingxinliu@ieee.org



Visit IMIC Lab