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

Mingxin Liu[1], Chengfei Cai[2], Anwen Lu[1], Pengbo Xu[3], Jun Li[1], Jinze Li[1], Depin Chen[1], and Jun Xu[1(✉)]

[1] Jiangsu Key Laboratory of Intelligent Medical Image Computing, School of Artificial Intelligence, Nanjing University of Information Science and Technology, Nanjing, China
jxu@nuist.edu.cn; chengfeicai@tzu.edu.cn
[2] Jiangsu Key Laboratory of Intelligent Drug Screening and Repositioning, College of Information Engineering, Taizhou University, Taizhou, China
[3] College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, China

Abstract. In computational pathology (CPath), developing omni-modal self-supervised learning (SSL) models that integrate histology, genomics, and clinical reports enables transferable representation learning for whole slide images (WSIs). Existing approaches implicitly force heterogeneous modalities into a uniform latent space by contrastive alignment, causing modality collapse where unique, synergistic diagnostic signals (termed as Φ) are discarded in favor of trivial redundancy. We hypothesize that the strongest task-agnostic SSL training signal stems from distilling the synergistic interactions over merely aligning shared redundancy. To this end, we introduce **Φ-Omni**, a synergistic information disentanglement framework grounded in Partial Information Decomposition (PID) theory for slide representation learning. Unlike standard contrastive approaches, Φ-Omni employs a Synergistic Information Bottleneck (SIB) regulated by the proposed ΦID objective, which explicitly suppresses marginal redundancy while maximizing irreducible synergy, thereby distilling high-order cross-modal interactions. Following pretraining on breast (n=1031) and lung (n=919) cohorts, Φ-Omni demonstrates superior few-shot performance across five independent external datasets spanning eight tasks compared to supervised and SSL baselines. Source code is available here.

Keywords: Computational Pathology · Slide Representation Learning · Multimodal Learning · Partial Information Decomposition.

1 Introduction

In recent years, self-supervised learning (SSL) has driven transformative progress in computational pathology (CPath), establishing itself as a foundational framework for CPath models to deliver superior performance in diagnostic, prognostic, and treatment response prediction tasks [1,12,13,14,15]. However, since the

whole slide images (WSIs) often exceed 150,000×150,000 pixels, most CPath approaches rely on a *divide-and-conquer* pipeline: (1) tessellating WSIs into small patches, (2) extracting patch features using a frozen pretrained SSL model, and (3) aggregating patch embeddings through Multiple Instance Learning (MIL) for prediction. In addition, SSL can generate universal *slide embeddings* from patch features by pretraining the aggregator further, enabling zero- or few-shot transfer across diverse downstream tasks without task-specific fine-tuning [7,17,25,8].

However, most slide representation learning methods remain unimodal, where neglecting complementary information from pathology reports and genomic profiles that is essential for robust and generalizable representations [25]. Therefore, recent efforts focus on multimodal pretraining. Early attempts explored multiple visual "views": MADELEINE [7] learns stain-invariant representations via cross-stain contrastive alignment, and BioX-CPath [3] integrates spatial and semantic features across stains with a graph architecture. Subsequent studies incorporate genomics: TANGLE [6] develops a transcriptomics-guided slide representation learning method using contrastive learning, while MIRROR [23] jointly optimizes shared representation alignment and modality-specific feature preservation.

Nevertheless, these approaches face two critical limitations: (1) pathology reports remain under-utilized despite providing expert diagnostic insights [15]. (2) prevalent reliance on contrastive alignment enforces redundancy, contradicting clinical intuition where diagnosis stems from the cross-modal synergy rather than their intersection alone [11]. This rigid alignment overlooks synergistic information, the emergent insights that are absent in any single modality yet solely accessible via joint multimodal observation, thus preventing the model from encoding the holistic representation essential for robust generalization [22].

To address these issues, we propose Φ-Omni, a novel SSL framework for omni-modal slide representation learning via synergistic information disentanglement (see Fig. 1). It firstly encodes omni-modal data into embeddings using domain-specific encoders, then these embeddings are further integrated and sent into the proposed Synergistic Information Bottleneck (SIB) module to distill high-order cross-modal interaction. Inspired by [19,22], we design a Synergistic Information Disentanglement objective (ΦID) with a Gaussian Canonical Projector (GCP) to regularize the omni-modal latent space by maximizing the synergistic interactions and minimizing the shared redundancy. The contributions are as follows:

- We propose Φ-Omni, an SSL framework for *slide representation learning* replacing redundancy alignment with synergistic disentanglement, which distills omni-modal insights into the slide encoder for robust unimodal inference.
- We introduce the SIB module and GCP with ΦID objective, which explicitly maximizes synergistic information while minimizing shared redundant noise to structure a discriminative omni-modal latent space.
- We validate Φ-Omni across eight diagnostic tasks on five public independent cohorts, demonstrating superior few-shot generalizabillity over state-of-the-art supervised MIL and SSL baselines.

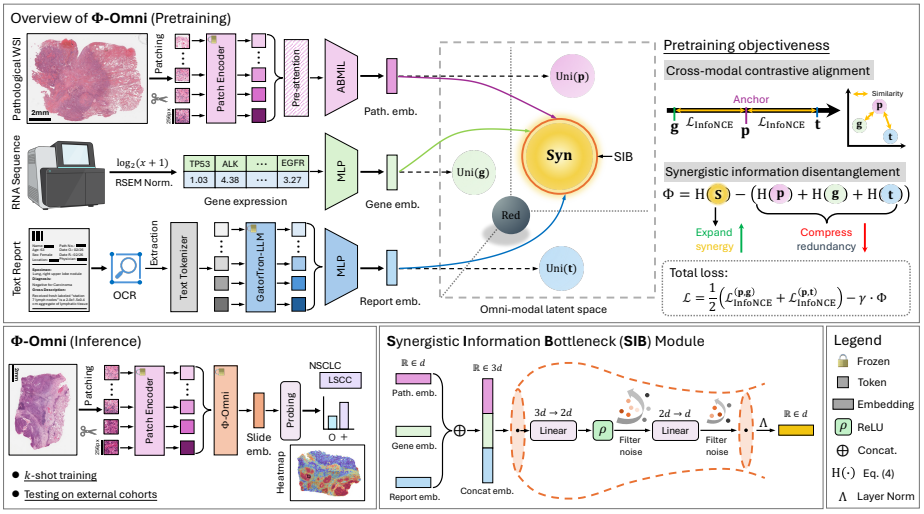

Fig. 1. The proposed Φ-Omni framework. Omni-modal inputs are embedded via domain-specific encoders then fused by Synergistic Information Bottleneck (SIB) module into a compact joint representation. The ΦID objective regularizes the omni-modal latent space, explicitly expanding emergent synergy while compressing redundancy. For inference, the frozen Φ-Omni slide encoder enables robust unimodal few-shot adaption.

2 Method

2.1 Omni-modal Feature Encoding

Let $\mathcal{D} = \{(\mathbf{X}_p^{(i)}, \mathbf{X}_g^{(i)}, \mathbf{X}_t^{(i)})\}_{i=1}^{N}$ denote the omni-modal dataset containing pathology WSIs, genomic profiles, and text reports for N patients. We employ domain-specific encoders to project these heterogeneous data into a triplet $\{\mathbf{p}_i, \mathbf{g}_i, \mathbf{t}_i\}$.

Pathological slide encoder. Given a WSI $\mathbf{X}_p^{(i)} \in \mathbb{R}^{d_x \times d_y \times 3}$ for the i^{th} patient, we tessellate it into 256×256 patches at 20× using TRIDENT [29] and extract it into patch embeddings using UNIv2 [1] as $\mathbf{P}_i \in \mathbb{R}^{\mathrm{D}_p \times 1536}$. These embeddings are then aggregated by a slide encoder f^{MIL}, implemented as an ABMIL [5] with a *pre-attention* layer [7], mapping the patch set into a slide embedding $\mathbf{p}_i \in \mathbb{R}^d$.

Transcriptomics encoder. We utilized paired Bulk RNA-seq data from UCSC Xena after log2(x+1) and RSEM transformed [4]. We select the genes associated with 50 MsigDB [10] Hallmark biological pathways to construct an input profile $\mathbf{G}_i \in \mathbb{R}^{\mathrm{D}_g}(\mathrm{D}_g \approx 5000)$, then encode it into a transcriptomics embedding $\mathbf{g}_i \in \mathbb{R}^d$ through a two-layer multilayer perceptron (MLP).

Text report encoder. We follow [20] to process pathology reports via an OCR-based pipeline with document image preprocessing (deskewing, noise reduction, binarization) and text cleaning (removing diagnostic labels). We encode the extracted text into semantic features $\mathbf{T}_i \in \mathbb{R}^{1024}$ via pre-trained GatorTron [28], and further project it into a text report embedding $\mathbf{t}_i \in \mathbb{R}^d$ via a two-layer MLP.

2.2 Synergistic Information Bottleneck Module

A key challenge in multimodal learning lies in the *modality gap*: naive integration propagates noise and enables dominant modality to suppress weaker ones. To address this, we propose the Synergistic Information Bottleneck (SIB) module to distill high-order cross-modal interactions. Structurally, SIB serves as an inductive bottleneck that maps the concatenated unimodal embeddings $\mathbf{X}_{\text{joint}} \in \mathbb{R}^{3d}$ to a compact joint representation $\mathbf{Z}_{\text{mm}} \in \mathbb{R}^{d}$ via linear projection layers $\text{Proj}(\cdot)$:

$$\mathbf{Z}_{\text{mm}} = \mathcal{F}_{\text{SIB}}\big(\mathbf{X}_{\text{joint}}\big) \triangleq \text{LN}\big(\text{Proj}_2\left(\text{ReLU}(\text{Proj}_1[\mathbf{p}_i \oplus \mathbf{g}_i \oplus \mathbf{t}_i])\right)\big) \tag{1}$$

where $\oplus$ and LN denote concatenation and Layer Normalization. By creating a dimensionality bottleneck ($\text{Proj}_1 \to \text{Proj}_2 : 3d \to 2d \to d$), SIB constrains the capacity of the multimodal transition, effectively filtering out modality-specific noise while capturing the most informative synergistic diagnostic signals.

2.3 Synergistic Information Disentanglement (ΦID)

Mutual information. The core objective of multimodal learning is to maximize the dependency between input modalities $\mathbf{X}$ and the fused feature $\mathbf{Z}$, typically measured via Mutual Information (MI) $\mathcal{I}(\mathbf{X}; \mathbf{Z})$. However, maximizing standard MI is insufficient for heterogeneous data due to its composite nature. According to Partial Information Decomposition (PID) theory [19], information from sources $\mathbf{X}$ about a target $\mathbf{Z}$ can be decomposed into three distinct atoms:
(1) **Redundancy (Red)**: Information shared across all modalities.
(2) **Uniqueness (Uni)**: Information provided solely by a specific modality $\mathbf{X}_m$.
(3) **Synergy (Syn)**: Emergent information is available *only* when observing multiple modalities jointly and is absent in any single source.
Formally, the total mutual information in this work satisfies:

$$\mathcal{I}(\mathbf{X}; \mathbf{Z}) = \text{Red}\left(\mathbf{X}; \mathbf{Z}\right) + \textstyle\sum_{m \in \{\mathbf{p},\mathbf{g},\mathbf{t}\}} \text{Uni}\left(\mathbf{X}_m; \mathbf{Z}\right) + \text{Syn}\left(\mathbf{X}; \mathbf{Z}\right) \tag{2}$$

standard fusion methods often maximize **Red**, as it represents the "least common denominator" features while causing modality collapse. In contrast, clinical diagnosis relies on **Syn** to integrate complementary cues for a holistic conclusion.
ΦID objective. To promote the emergence of **Syn** and mitigates the subspace collapse of **Red**, We denote omni-modal synergistic interaction as Φ, computed via the "*whole-minus-sum*" residual in PID to isolate multimodal information:

$$\Phi\left(\{\mathbf{p},\mathbf{g},\mathbf{t}\} \to \mathbf{Z}_{\text{mm}}\right) \triangleq \underbrace{\mathcal{I}\left(\mathbf{X}_{\text{joint}}; \mathbf{Z}_{\text{mm}}\right)}_{\text{Total correlation}} - \sum_{m \in \{\mathbf{p},\mathbf{g},\mathbf{t}\}} \underbrace{\mathcal{I}\left(\mathbf{X}_m; \mathbf{Z}_{\text{mm}}\right)}_{\text{Marginal redundancy}} \tag{3}$$

where maximizing Φ forces the model to capture high-order **Syn** interactions while minimizing the marginal **Red** that can be inferred from single modality.
Gaussian Canonical Projector. To ensure computational tractability for estimating MI in high dimensions, we design Gaussian Canonical Projector (GCP)

inspired by [21]. GCP leverages Layer Normalization to explicitly map heterogeneous embeddings $\mathbf{h} \in \{\mathbf{X}_{\text{joint}}, \mathbf{X}_m, \mathbf{Z}_{\text{mm}}\}$ onto a smooth Gaussian latent space for differential entropy estimation, enables $\mathrm{H}(\mathbf{Z})$ to be estimated via a parametric variational Gaussian approximation with a closed-form log-determinant:

$$\mathbf{Z} = \mathbf{\Pi}_{\text{GCP}}(\mathbf{h}) \triangleq \text{Sigmoid}\Big(\text{LN}\big(\mathbf{W} \cdot \mathbf{h}\big)\Big), \quad \mathrm{H}(\mathbf{Z}) = \frac{1}{2}\log\det\left(\Sigma_{\mathbf{Z}}\right) \tag{4}$$

where $\frac{1}{2}\log\det\left(\Sigma_{\mathbf{Z}}\right)$ captures feature "volume", maximized to expand information and prevent collapse. $\Sigma_{\mathbf{Z}} = \frac{1}{\mathrm{B}}\mathbf{Z}^\top\mathbf{Z} + \epsilon\mathbf{I}$ is the empirical covariance matrix for batch size B, $\mathbf{I}$ is the identity matrix with $\epsilon = 10^{-5}$ for numerical stability.

2.4 Pretraining Objective

Cross-modal contrastive alignment. We align the latent space via a symmetric InfoNCE objective [2]. Defining the directional pairwise loss $\ell(\mathbf{u}, \mathbf{v})$ between a query modality $\mathbf{u}$ and a key modality $\mathbf{v}$, the omni-modal symmetric contrastive alignment objective is calculated as:

$$\mathcal{L}_{\text{SymCL}} = \frac{1}{2}\sum_{k\in\{\mathbf{g},\mathbf{t}\}}\Big(\ell(\mathbf{p},\mathbf{k}) + \ell(\mathbf{k},\mathbf{p})\Big), \ell(\mathbf{u},\mathbf{v}) = -\frac{1}{\mathrm{B}}\sum_{i=1}^{\mathrm{B}}\log\frac{e^{\tau\mathbf{u}_i^\top\mathbf{v}_i}}{\sum_{j=1}^{\mathrm{M}}e^{\tau\mathbf{u}_j^\top\mathbf{v}_j}} \tag{5}$$

where τ is the temperature parameter, and the summation over $k = \{\mathbf{g}, \mathbf{t}\}$ aligns the pathology anchor (central modality) $\mathbf{p}$ with both transcriptomics $\mathbf{g}$ and report embeddings $\mathbf{t}$, skipping auxiliary $\mathbf{g}$-$\mathbf{t}$ alignment for efficiency.

Synergistic information disentanglement. To distill **Syn**, we minimize the Total Correlation (TC) of projected margins as a parametric variational proxy for Eq. (3). This approximation compresses redundancy, forcing the joint embedding to capture synergistic interactions. The loss for ΦID objective is formulated as:

$$\mathcal{L}_{\Phi\text{ID}} = -\Phi = \textstyle\sum_{m\in\{\mathbf{p},\mathbf{g},\mathbf{t}\}}\mathrm{H}(\mathbf{Z}_m) - \mathrm{H}(\mathbf{Z}_{\text{joint}}) \tag{6}$$

where $\mathbf{Z}_* = \mathbf{\Pi}_{\text{GCP}}(\mathbf{X}_*)$. Specifically, this objective imposes a dual constraint: (1) it compresses the marginal entropies to suppress unimodal redundancy, while (2) expanding the joint entropy to capture synergistic cross-modal interactions.

Complementary objective. Overall, we train Φ-Omni using a composite loss: $\mathcal{L} = \mathcal{L}_{\text{SymCL}} + \gamma\mathcal{L}_{\Phi\text{ID}}$, where the former ensures foundational semantic alignment and the latter acts as a regularizer that penalizes trivial redundancy to encourage the discovery of complementary synergistic features. We set $\gamma = 0.2$ in this work after hyperparameter searching over $\{0.01, 0.1, 0.2, 0.5, 0.8, 1.0\}$.

3 Experimental Setups and Results

3.1 Study Design

To validate Φ-Omni, we established two omni-modal pretraining cohorts by filtering TCGA [1] cases with strictly triplet-aligned modalities (pathological WSI, transcriptomics, pathology reports). For inference, only WSIs are required.

[1] https://portal.gdc.cancer.gov/

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 (↑)				CPTAC-NSCLC (↑)			
		k=1	k=5	k=10	k=25	k=1	k=5	k=10	k=25
MIL	ABMIL [5]	55.9 ± 2.8	64.8 ± 2.0	75.4 ± 2.6	82.3 ± 0.8	72.2 ± 12.4	91.9 ± 5.2	95.7 ± 2.7	98.3 ± 1.0
MIL	CLAM [16]	55.5 ± 3.0	65.4 ± 3.2	74.9 ± 2.0	82.3 ± 1.2	74.3 ± 14.1	93.6 ± 3.1	96.5 ± 2.3	98.5 ± 0.7
MIL	TransMIL [18]	57.1 ± 2.7	68.9 ± 2.2	75.6 ± 1.2	82.3 ± 1.9	71.7 ± 12.2	91.8 ± 3.0	95.2 ± 1.4	98.6 ± 0.5
MIL	ILRA [26]	55.0 ± 1.8	62.5 ± 2.8	70.0 ± 1.6	79.0 ± 1.7	66.7 ± 15.9	87.1 ± 2.5	92.4 ± 1.7	97.0 ± 1.1
Linear probe	CTransPath* [24]	56.7 ± 2.8	63.3 ± 2.9	64.7 ± 1.4	68.9 ± 1.6	71.3 ± 11.1	88.0 ± 4.9	92.3 ± 2.1	96.0 ± 1.1
Linear probe	UNIv2* [1]	55.8 ± 3.2	66.0 ± 2.0	70.6 ± 1.7	76.4 ± 1.5	72.9 ± 15.3	91.4 ± 4.2	95.3 ± 1.7	97.5 ± 1.0
Linear probe	CONCH* [15]	58.4 ± 8.4	69.5 ± 2.8	73.4 ± 2.1	79.0 ± 1.7	**83.1 ± 9.3**	95.8 ± 2.0	97.0 ± 0.7	98.5 ± 0.5
Linear probe	mSTAR* [27]	56.1 ± 3.0	66.0 ± 2.3	70.2 ± 2.0	75.4 ± 1.3	74.9 ± 11.9	92.1 ± 4.2	95.5 ± 1.7	97.8 ± 0.8
Linear probe	GPFM* [17]	56.4 ± 2.7	66.2 ± 2.8	70.9 ± 2.0	75.8 ± 1.3	73.2 ± 10.8	90.3 ± 4.9	95.3 ± 1.9	97.8 ± 1.0
Linear probe	CHIEF [25]	64.2 ± 3.2	73.6 ± 2.1	77.5 ± 1.4	81.9 ± 1.1	72.7 ± 10.9	91.0 ± 4.7	94.6 ± 1.6	97.4 ± 0.7
Linear probe	TANGLE [6]	62.2 ± 2.3	74.1 ± 2.4	78.1 ± 1.2	82.4 ± 0.8	79.3 ± 10.2	95.3 ± 4.3	96.9 ± 2.1	98.8 ± 0.2
Linear probe	MADELEINE [7]	61.3 ± 4.5	73.4 ± 2.4	76.7 ± 2.0	82.3 ± 1.3	81.3 ± 9.6	94.7 ± 2.7	96.6 ± 1.1	98.2 ± 0.6
Linear probe	Φ-Omni (Ours)	**65.2 ± 2.6**	**76.5 ± 2.2**	**79.9 ± 1.0**	**83.3 ± 0.8**	81.7 ± 11.3	**97.5 ± 1.6**	**98.8 ± 0.5**	**99.4 ± 0.2**

Breast. For breast **pretraining**, we used n=1,031 omni-modal primary breast cases from TCGA Breast Invasive Carcinoma (BRCA) cohort. Downstream evaluation was conducted on two independent cohorts: (1) BRACS dataset [2] (n=547) for 7-way breast cancer fine-grained subtyping and (2) CPTAC-BRCA [3] dataset (n=112) for predicting the molecular mutation status of TP53 and PIK3CA.
Lung. For lung **pretraining**, we used n=919 omni-modal primary lung cases from TCGA Non-Small Cell Lung Cancer (NSCLC) cohort. We performed downstream evaluation on CPTAC-NSCLC for: (1) lung cancer subtyping (LUAD *vs.* LSCC, n=604), molecular mutation prediction of (2) STK11/TP53 on CPTAC-LUAD (n=324), (3) ARID1A/KEAP1 on CPTAC-LSCC (n=304).

3.2 Evaluation and Implementation Details

Few-shot classification. Following [6,7], we benchmark Φ-Omni against SSL models via k-shot linear probing (k=1, 5, 10, 25) via a logistic regression classifier (L2 penalty C=1.0; max iterations=10^4) from scikit-learn. All experiments are repeated ten times by randomly sampling k examples per class during training with different seeds, reporting the mean and standard deviation of macro-AUC.
Implementation details. Φ-Omni was trained for 100 epochs (5 for warmup) using AdamW with cosine decay (10^{-4} to 10^{-8}) and batch size 128. We sampled 2,048 fixed patches per slide with random oversampling applied to slides with fewer patches. All MIL models were trained from scratch with batch size of 1.
Baselines. We benchmark Φ-Omni against four MIL models based on UNIv2 [1] encoder: ABMIL [5], CLAM [16], TransMIL [18], and ILRA [26]. We also compare with SSL models via linear probing, including patch-level SSLs: CTransPath [24], UNIv2 [1], CONCH [15], mSTAR [27], and GPFM [17] using mean-pool embeddings (noted $*$), and slide-levels: CHIEF [25], TANGLE [6], and MADELEINE [7].

[2] https://www.bracs.icar.cnr.it/
[3] https://proteomic.datacommons.cancer.gov/pdc/

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 (↑) PIK3CA	BRCA (↑) TP53	LUAD (↑) STK11	LUAD (↑) TP53	LSCC (↑) ARID1A	LSCC (↑) KEAP1	Avg. (↑)
MIL	ABMIL [5]	ICML'2018	61.2 ± 8.7	76.1 ± 4.6	90.8 ± 5.1	76.3 ± 5.0	82.2 ± 5.0	84.5 ± 3.8	78.5
MIL	CLAM [16]	Nat. BME.'2021	60.5 ± 8.7	75.5 ± 5.2	92.1 ± 4.4	78.3 ± 3.9	81.0 ± 5.1	84.4 ± 5.0	78.6
MIL	TransMIL [18]	NeurIPS'2021	59.4 ± 7.6	78.6 ± 3.5	91.3 ± 3.3	78.0 ± 4.8	80.9 ± 3.6	83.9 ± 5.2	78.7
MIL	ILRA [26]	ICLR'2023	61.4 ± 7.9	74.5 ± 5.9	91.7 ± 4.3	77.8 ± 4.5	81.3 ± 4.0	87.1 ± 6.3	79.0
Linear probe	CTransPath* [24]	MedIA'2022	62.1 ± 4.8	73.1 ± 4.8	85.3 ± 3.4	70.8 ± 2.7	75.2 ± 7.5	82.4 ± 3.5	74.8
Linear probe	UNIv2* [1]	Nat. Med.'2024	59.6 ± 5.2	77.2 ± 5.2	92.0 ± 2.4	77.7 ± 3.0	80.3 ± 7.3	87.6 ± 5.0	79.1
Linear probe	CONCH* [15]	Nat. Med.'2024	56.2 ± 5.3	77.2 ± 5.0	89.2 ± 3.3	68.9 ± 4.1	73.2 ± 9.8	78.6 ± 5.2	73.9
Linear probe	mSTAR* [27]	Nat. Comm.'2025	60.0 ± 4.8	78.1 ± 4.6	91.5 ± 2.6	77.2 ± 2.5	79.2 ± 7.0	87.5 ± 4.6	78.9
Linear probe	GPFM* [17]	Nat. BME.'2025	59.9 ± 5.3	77.5 ± 4.2	91.9 ± 2.9	74.9 ± 3.0	78.8 ± 7.8	85.5 ± 4.5	78.1
Linear probe	CHIEF [25]	Nature'2024	63.8 ± 6.2	78.1 ± 4.1	89.5 ± 3.2	74.0 ± 2.5	78.8 ± 7.3	81.7 ± 3.4	77.7
Linear probe	TANGLE [6]	CVPR'2024	60.1 ± 6.2	**81.0 ± 3.5**	91.8 ± 2.7	79.7 ± 3.5	82.8 ± 5.3	85.6 ± 4.6	80.2
Linear probe	MADELEINE [7]	ECCV'2024	61.4 ± 5.9	79.3 ± 4.6	92.1 ± 2.6	72.7 ± 2.5	75.9 ± 8.2	83.1 ± 6.4	77.4
Linear probe	Φ-Omni (Ours)	MICCAI'2026	**64.0 ± 6.1**	80.0 ± 3.6	**93.3 ± 1.7**	**82.4 ± 2.5**	**85.1 ± 5.1**	**88.8 ± 5.2**	**82.3**

3.3 Few-shot Classification Results

Φ-Omni *vs.* UNIv2 *vs.* TANGLE. Φ-Omni exceeds UNIv2 (pathology) and TANGLE (pathology+genomic) across all tasks (Table 1,2), highlights the value of clinical (report) and biological (genomic) "views" from multimodal pretraining.
Φ-Omni *vs.* MILs. Φ-Omni surpasses all MILs in 8/8 tasks, achieving +4.3%, +1.2%, and +1.1% gains compared to ABMIL (k=25) for molecular prediction, breast and lung subtyping, respectively. Notably, Φ-Omni employs linear probing only, whereas MIL models are trained from scratch in standard few-shot settings.
Φ-Omni *vs.* SSLs. Φ-Omni outperforms all SSL models on most tasks, often by notable margins, *e.g.*, +3.2% over TANGLE (BRACS, breast cancer subtyping, k=5) and +5.7% over UNIv2 (CPTAC-LSCC, ARID1A mutation prediction, k=25). Notably, this advantage holds consistently across all k values and tasks, underscoring the excellent performance and robust generalization of Φ-Omni.

3.4 Ablation Study

Architecture ablation. We explore two key architectural features as shown in Fig. 2. (a) the network for slide encoding: we compare ABMIL w/ *pre-attention* with MLP, TransMIL, and ABMIL, our choice yields notable improvements. (b) the module for multimodal feature fusion: we evaluate Trilinear [9], Concat, and our SIB module, where SIB outperforms alternatives steadily for all six tasks.
Loss ablation. We perform a thorough ablation of Φ-Omni loss function by re-training models with L1 objective, Mean-Squared Error (MSE), InfoNCE, symmetric contrastive objective (SymCL), and combining SymCL and ΦID (Ours). Fig. 2 (c) shows that our ΦID objective is indispensable: SymCL alone achieves strong gains over L1/MSE/InfoNCE, but combining it with ΦID unlocks synergistic improvements across all six tasks (+4.4% on BRACS-Subtyping and +7.2% on BRCA-PIK3CA), proving that explicit synergy disentanglement (not just contrastive alignment) is critical for maximizing multimodal diagnostic value.

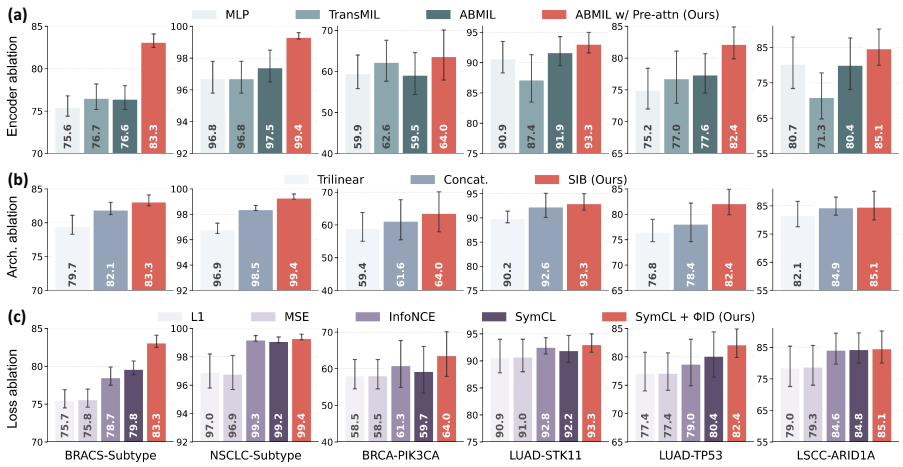

Fig. 2. Ablation study results on architecture and loss function across six tasks.

3.5 Interpretability Analysis

Attention heatmaps. We generate attention heatmaps of Φ-Omni pretrained on lung/breast cancer (Fig. 3 (a)), which reveals spatial correspondence between high-attention regions and tumor areas, demonstrating task-agnostic supervised localization of pathological regions through multimodal pretraining.

Omni-modal disentanglement. To investigate the structural properties of the learned latent space, we visualize omni-modal embeddings extracted from Φ-Omni pretrained on lung cancer via t-SNE (Fig. 3 (b)). The visualization exhibits three well-separated clusters, indicating that Φ-Omni effectively learns discriminative representations for each modality, preventing the modality collapse. Notably, the explicit connections between paired samples ("Cross-modal Synergy") confirm that Φ-Omni preserves strong semantic alignment across heterogeneous modalities while maintaining the intrinsic biological and clinical correlations.

4 Conclusion

In this paper, we introduce Φ-Omni, a novel SSL framework that advances omni-modal slide representation learning by shifting the focus from redundancy-driven contrastive alignment to synergistic information disentanglement. Derived from Partial Information Decomposition (PID) theory, our proposed SIB module and ΦID objective provide a mathematically principled mechanism to extract high-order diagnostic insights that are absent in any single modality. Extensive experiments across six benchmarks demonstrate that Φ-Omni not only achieves state-of-the-art few-shot performance but maintains biologically grounded interpretability. This work highlights the power of distilling synergistic interaction from omni-modal histology, genomics, and pathology reports, paving the way for data-efficient and robust multimodal frameworks in computational pathology.

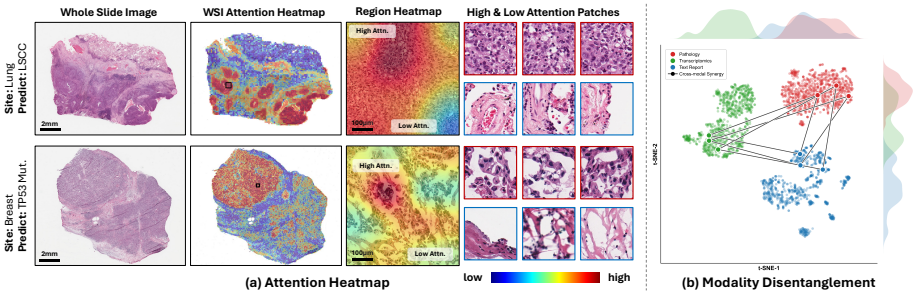

Fig. 3. Interpretability analysis of Φ-Omni. (a) Attention heatmaps of the frozen ABMIL slide encoder pretrained with Φ-Omni overlaid on randomly chosen WSIs from lung and breast cohort. Red/blue denotes high/low attention regions with corresponding pathological patch details. **(b)** t-SNE visualization: Φ-Omni disentangles cross-modal representations (pathology, transcriptomics, text reports) into distinct clusters.

Acknowledgments. This study is funded by National Key R&D Program of China (No. 2023YFC3402800), National Natural Science Foundation of China (Nos. 82441029, 62171230, 62101365, 92159301, 62301263, 62301265, 62302228, 82302291, 82302352, 62401272), Jiangsu Provincial Department of Science and Technology's major project on frontier-leading basic research in technology (No. BK2023200), and Taizhou University Research Start-up Fund for High-Level Talents (No. TZXYQD2025B101).

Disclosure of Interests. The authors declare no relevant competing interests.

References

1. Chen, R.J., Ding, T., Lu, M.Y., Williamson, D.F., Jaume, G., Song, A.H., Chen, B., Zhang, A., Shao, D., Shaban, M., et al.: Towards a general-purpose foundation model for computational pathology. Nature medicine **30**(3), 850–862 (2024)
2. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International conference on machine learning. pp. 1597–1607. PmLR (2020)
3. Gallagher-Syed, A., Senior, H., Alwazzan, O., Pontarini, E., Bombardieri, M., Pitzalis, C., Lewis, M.J., Barnes, M.R., Rossi, L., Slabaugh, G.: Biox-cpath: Biologically-driven explainable diagnostics for multistain ihc computational pathology. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 10372–10383 (2025)
4. Goldman, M.J., Craft, B., Hastie, M., Repečka, K., McDade, F., Kamath, A., Banerjee, A., Luo, Y., Rogers, D., Brooks, A.N., Zhu, J., Haussler, D.: Visualizing and interpreting cancer genomics data via the xena platform. Nature biotechnology **38**(6), 675–678 (2020)
5. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International conference on machine learning. pp. 2127–2136. PMLR (2018)
6. Jaume, G., Oldenburg, L., Vaidya, A., Chen, R.J., Williamson, D.F., Peeters, T., Song, A.H., Mahmood, F.: Transcriptomics-guided slide representation learning in

computational pathology. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9632–9644 (2024)
7. Jaume, G., Vaidya, A., Zhang, A., H. Song, A., J. Chen, R., Sahai, S., Mo, D., Madrigal, E., Phi Le, L., Mahmood, F.: Multistain pretraining for slide representation learning in pathology. In: European Conference on Computer Vision. pp. 19–37. Springer (2024)
8. Lazard, T., Lerousseau, M., Decencière, E., Walter, T.: Giga-ssl: Self-supervised learning for gigapixel images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4305–4314 (2023)
9. Li, H., Yang, F., Xing, X., Zhao, Y., Zhang, J., Liu, Y., Han, M., Huang, J., Wang, L., Yao, J.: Multi-modal multi-instance learning using weakly correlated histopathological images and tabular clinical information. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 529–539. Springer (2021)
10. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., Tamayo, P.: The molecular signatures database hallmark gene set collection. Cell systems **1**(6), 417–425 (2015)
11. Liu, M., Cai, C., Li, J., Xu, P., Li, J., Ma, J., Xu, J.: Murrenet: Modeling holistic multimodal interactions between histopathology and genomic profiles for survival prediction. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 396–406. Springer (2025)
12. Liu, M., Liu, Y., Cui, H., Li, C., Ma, J.: Mgct: Mutual-guided cross-modality transformer for survival outcome prediction using integrative histopathology-genomic features. In: 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 1306–1312. IEEE (2023)
13. Liu, M., Liu, Y., Xu, P., Cui, H., Ke, J., Ma, J.: Exploiting geometric features via hierarchical graph pyramid transformer for cancer diagnosis using histopathological images. IEEE Transactions on Medical Imaging (2024)
14. Liu, M., Liu, Y., Xu, P., Ma, J.: Unleashing the infinity power of geometry: A novel geometry-aware transformer (goat) for whole slide histopathology image analysis. In: 2024 IEEE International Symposium on Biomedical Imaging (ISBI). pp. 1–5. IEEE (2024)
15. Lu, M.Y., Chen, B., Williamson, D.F., Chen, R.J., Liang, I., Ding, T., Jaume, G., Odintsov, I., Le, L.P., Gerber, G., et al.: A visual-language foundation model for computational pathology. Nature medicine **30**(3), 863–874 (2024)
16. Lu, M.Y., Williamson, D.F., Chen, T.Y., Chen, R.J., Barbieri, M., Mahmood, F.: Data-efficient and weakly supervised computational pathology on whole-slide images. Nature biomedical engineering **5**(6), 555–570 (2021)
17. Ma, J., Guo, Z., Zhou, F., Wang, Y., Xu, Y., Li, J., Yan, F., Cai, Y., Zhu, Z., Jin, C., et al.: A generalizable pathology foundation model using a unified knowledge distillation pretraining framework. Nature Biomedical Engineering pp. 1–20 (2025)
18. Shao, Z., Bian, H., Chen, Y., Wang, Y., Zhang, J., Ji, X., et al.: Transmil: Transformer based correlated multiple instance learning for whole slide image classification. Advances in neural information processing systems **34**, 2136–2147 (2021)
19. Tokui, S., Sato, I.: Disentanglement analysis with partial information decomposition. arXiv preprint arXiv:2108.13753 (2021)
20. Tripathi, A., Waqas, A., Schabath, M.B., Yilmaz, Y., Rasool, G.: Honeybee: enabling scalable multimodal ai in oncology through foundation model-driven embeddings. npj Digital Medicine **8**(1), 622 (2025)

21. Venkatesh, P., Bennett, C., Gale, S., Ramirez, T., Heller, G., Durand, S., Olsen, S., Mihalas, S.: Gaussian partial information decomposition: Bias correction and application to high-dimensional data. Advances in Neural Information Processing Systems **36**, 74602–74635 (2023)
22. Ver Steeg, G., Brekelmans, R., Harutyunyan, H., Galstyan, A.: Disentangled representations via synergy minimization. In: 2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton). pp. 180–187. IEEE (2017)
23. Wang, T., Fan, J., Zhang, D., Liu, D., Xia, Y., Huang, H., Cai, W.: Mirror: Multimodal pathological self-supervised representation learning via modality alignment and retention. arXiv preprint arXiv:2503.00374 (2025)
24. Wang, X., Yang, S., Zhang, J., Wang, M., Zhang, J., Yang, W., Huang, J., Han, X.: Transformer-based unsupervised contrastive learning for histopathological image classification. Medical image analysis **81**, 102559 (2022)
25. Wang, X., Zhao, J., Marostica, E., Yuan, W., Jin, J., Zhang, J., Li, R., Tang, H., Wang, K., Li, Y., et al.: A pathology foundation model for cancer diagnosis and prognosis prediction. Nature **634**(8035), 970–978 (2024)
26. Xiang, J., Zhang, J.: Exploring low-rank property in multiple instance learning for whole slide image classification. In: The Eleventh International Conference on Learning Representations (2023)
27. Xu, Y., Wang, Y., Zhou, F., Ma, J., Jin, C., Yang, S., Li, J., Zhang, Z., Zhao, C., Zhou, H., et al.: A multimodal knowledge-enhanced whole-slide pathology foundation model. Nature Communications (2025)
28. Yang, X., Chen, A., PourNejatian, N., Shin, H.C., Smith, K.E., Parisien, C., Compas, C., Martin, C., Costa, A.B., Flores, M.G., et al.: A large language model for electronic health records. NPJ digital medicine **5**(1), 194 (2022)
29. Zhang, A., Jaume, G., Vaidya, A., Ding, T., Mahmood, F.: Accelerating data processing and benchmarking of ai models for pathology. arXiv preprint arXiv:2502.06750 (2025)