LEARNING INTRA-VIEW COMPLEMENTARITY FOR HIGH-DIMENSIONAL MULTI-VIEW CLUSTERING

Anonymous authors
Paper under double-blind review

ABSTRACT

Multi-view clustering mainly exploits complementarity across views, while often overlooking structural diversity within each individual view. This limitation becomes more pronounced in high-dimensional settings, where heterogeneous structural cues can coexist within a single view and may not be adequately captured by a single representation. We propose LIVC, a framework for Learning Intra-View Complementarity in high-dimensional multi-view clustering. LIVC first constructs raw, global-principal, and local-structural representations to expose complementary structures within each view. It then performs granularity-aligned cross-view interaction with learnable gating for selective information transfer. Finally, residual low-rank bilinear fusion captures second-order dependencies with controlled model complexity. A joint self-supervised objective preserves view-specific information while coordinating representations across granularities and views. Experiments on multiple high-dimensional multi-view datasets demonstrate the effectiveness of LIVC for clustering.

1 INTRODUCTION

Modern datasets are increasingly characterized by multiple heterogeneous views and high-dimensional feature spaces. Multi-omics data are a representative example, where each patient is described by multiple molecular profiles with thousands of features but relatively limited samples. Such a dimensionality–sample imbalance makes clustering difficult, as redundant and noisy features can obscure latent structures and destabilize representation learning. These challenges motivate multi-view clustering methods that can effectively exploit robust and complementary information from high-dimensional heterogeneous views.

To exploit complementary information across heterogeneous views, existing Multi-View Clustering (MVC) methods have mainly evolved along two directions. Representation-and-fusion approaches integrate view-specific features through shared latent representations (Xu et al., 2026), adaptive graph or anchor learning (Guo et al., 2026a; Tang et al., 2026), disentangled modeling (Xiao et al., 2026; Zhang & Che, 2024), and higher-order fusion (Kong et al., 2026; Guo et al., 2026b; Wang et al., 2026a), whereas contrastive and structure-aware approaches promote cross-view consistency through representation alignment (Yuan et al., 2026; Guo et al., 2024), structural constraints (Fei et al., 2025; Fu et al., 2024; Liu et al., 2024), pseudo-label supervision (Yu et al., 2025; Hu et al., 2025), and multi-level semantic learning (Chen et al., 2025; Wang et al., 2023; Bian et al., 2025). Despite their different formulations, both lines of research primarily organize information across views, with the individual view typically represented by a single feature space or structural form. This cross-view-centric perspective may be insufficient for high-dimensional data, where rich and heterogeneous structures can already coexist within each view.

More specifically, a high-dimensional view may exhibit heterogeneous and complementary structural patterns that are not equally preserved by a single representation. Relying on a single representation may therefore fail to capture these structures simultaneously, potentially discarding useful information before cross-view integration. This motivates explicitly modeling and coordinating diverse structural cues within each view. Once these intra-view structures are exposed, the next challenge is how to exploit them across views without destroying their structural roles. Representations at different granularities encode distinct semantics, making unrestricted cross-view interaction prone to mixing structurally mismatched information. Moreover, even structurally corresponding

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

information from other views may vary in relevance and reliability. Effective cross-view learning should therefore preserve granularity correspondence while selectively transferring informative signals.

To address these challenges, we propose LIVC, a framework for high-dimensional multi-view clustering that learns intra-view complementarity before cross-view integration. Rather than compressing each view into a single representation, LIVC preserves its raw information while deriving global-principal and local-neighborhood representations, and coordinates these complementary perspectives in a shared latent space. Building on these intra-view representations, LIVC performs granularity-aligned cross-view interaction, where structurally corresponding branches selectively exchange information through learnable gating, and subsequently consolidates the enhanced structural cues around the raw semantic anchor. Finally, LIVC introduces a residual refinement path and applies low-rank bilinear fusion to model second-order interactions between the enhanced and refined representations with controlled complexity. Our main contributions are summarized as follows:

- We introduce a within-view complementary representation strategy for high-dimensional multi-view clustering. Rather than compressing each view into a single embedding, we preserves the raw information as a semantic reference while deriving global-principal and local-neighborhood representations as complementary structural cues, enabling heterogeneous information within each view to be jointly exploited before cross-view integration.
- We develop a granularity-aligned selective cross-view enhancement mechanism that preserves the structural roles of different representations during information exchange. Only structurally corresponding branches interact across views, while learnable gating controls the contribution of transferred information and the enhanced structural cues are consolidated around the raw semantic anchor.
- We propose a residual low-rank bilinear fusion mechanism to model multiplicative dependencies beyond conventional additive fusion. By factorizing the bilinear interaction into a compact latent space while retaining a residual first-order path, the proposed fusion captures second-order interactions without constructing the quadratic full bilinear space.

2 Related Work

2.1 Multi-View Representation and Fusion Learning

MVC aims to exploit shared and complementary information across heterogeneous views. Recent methods have explored cross-view consistency through neural representation learning, adaptive fusion, graph modeling, and disentangled representations. For example, BONE (Xu et al., 2026) combines optimization priors with lightweight neural networks, while MISCC (Li et al., 2026b) jointly models intra-view similarity and cross-view correlation. DMCAR-MVC (Xiao et al., 2026) and GMAE (Zou et al., 2026) further separate shared and view-specific information to preserve representation diversity. Information-theoretic and tensor-based approaches, such as STMVC (Ren et al., 2026), ARIA (Kong et al., 2026), and TPSC (Wang et al., 2026b), model redundancy, low-rank structure, or higher-order dependencies across views. Multi-granularity anchor learning has also been explored to adapt representation scales across views. Despite these advances, most methods primarily organize information at the view level, leaving the heterogeneous structural cues contained within each individual high-dimensional view underexplored.

2.2 Contrastive and Structure-Aware Multi-View Clustering

Contrastive learning has become a widely adopted paradigm for unsupervised multi-view representation alignment. Existing studies improve cross-view learning from multiple perspectives. CDARL (Xi et al., 2025) combines contrastive and adversarial learning to reduce discrepancies among view-specific latent distributions, while PSGVC (Li et al., 2026a) introduces pseudo-label similarity graphs to provide clustering-aware supervision. ILASC (Yang & Tan, 2026) further aligns sample-level clustering structures across views, whereas DWCL (Yuan et al., 2026) selectively constructs reliable cross-view pairs to alleviate unreliable view combinations.

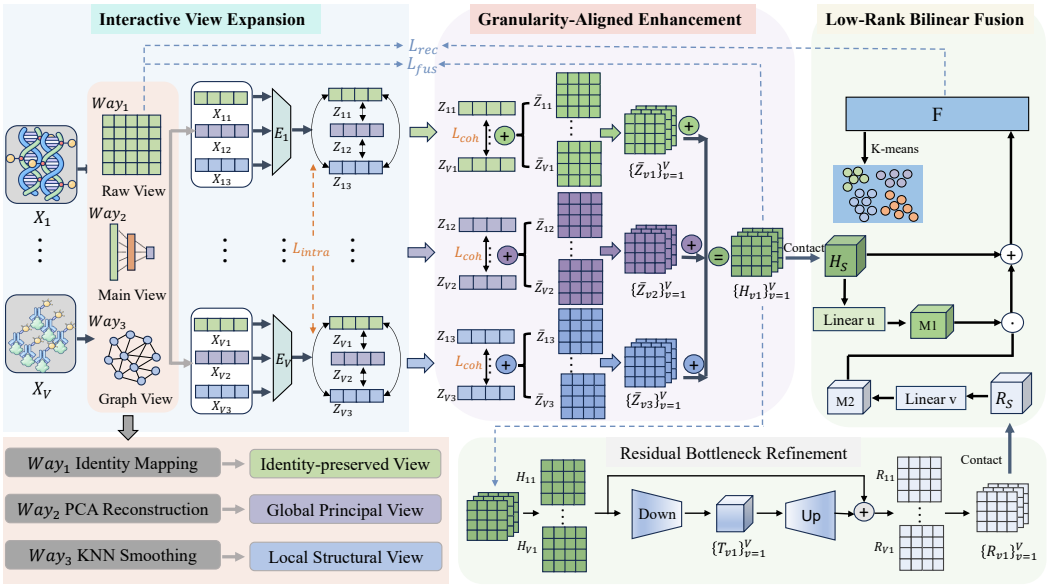


Figure 1: Overview of LIVC. The Interactive View Expansion module preserves raw information while deriving complementary global-principal and local-neighborhood representations. The Granularity-Aligned Enhancement module selectively exchanges information across views between structurally corresponding representations and consolidates them around the raw semantic anchor. The Low-Rank Bilinear Fusion module captures second-order interactions with controlled complexity to produce the unified representation for clustering.

Recent studies also incorporate uncertainty and structural information into contrastive learning. STCMC-UR (Hu et al., 2025) employs uncertainty to refine contrastive pairs, and DMvCGSA (Fei et al., 2025) performs both instance- and cluster-level contrastive learning with graph structure awareness. SERIES (Wang & Feng, 2024) explicitly models view-interaction structures, while SCMC (Fu et al., 2024) and PNCL-FDMC (Liu et al., 2024) preserve local geometric or neighborhood structures during representation alignment. Meanwhile, Deep MVC shows that excessive contrastive alignment may impair cluster separability. Despite these advances, most approaches primarily exploit structural correspondence *across views*, whereas structural complementarity among multiple granularities *within each individual view* remains less explored.

3 Methodology

We consider high-dimensional multi-view clustering with V views, where $\mathbf{X}^{(v)} \in \mathbb{R}^{N \times d_v}$ denotes the v-th view of N samples. Our goal is to learn a unified representation that preserves informative within-view structures while exploiting cross-view complementarity. To this end, we propose LIVC (Learning Intra-View Complementarity), which follows a within-view-before-cross-view strategy. For each view, LIVC preserves raw information as a semantic reference while deriving main and graph representations to capture global principal structure and local neighborhood relations. It then performs granularity-aligned cross-view enhancement with selective information transfer and consolidates the enhanced structural cues around the raw semantic anchor. Finally, residual bottleneck refinement and low-rank bilinear fusion capture second-order interactions with controlled complexity. LIVC consists of three modules: Interactive View Expansion, Granularity-Aligned Enhancement, and Low-Rank Bilinear Fusion, as illustrated in Figure 1.

3.1 Interactive View Expansion

A high-dimensional view contains complementary structural information at different granularities. To capture this diversity, we expand each view $\mathbf{X}_v$ into three sample-aligned branches: raw, main,

and graph. The raw branch preserves standardized original features:

$$\mathbf{X}_v^r = \mathrm{Norm}(\mathbf{X}_v), \tag{1}$$

where $\mathrm{Norm}(\cdot)$ denotes feature-wise standardization. The main and graph branches capture global principal structure and local neighborhood relations, respectively:

$$\mathbf{X}_v^m = \mathrm{PCA}_{q_v}^{-1}\left(\mathrm{PCA}_{q_v}(\mathbf{X}_v^r)\right), \qquad \mathbf{X}_v^g = \mathbf{D}_v^{-1}\mathbf{A}_v\mathbf{X}_v^r, \tag{2}$$

where q_v is the minimum number of principal components meeting a predefined cumulative explained-variance threshold, and $\mathbf{A}_v$ and $\mathbf{D}_v$ denote the symmetrized KNN adjacency matrix and its degree matrix.

The three branches are mapped into a common latent space through a shared view-specific encoder E_v:

$$\mathbf{Z}_v^b = E_v(\mathbf{X}_v^b), \qquad b \in \{r, m, g\}. \tag{3}$$

Parameter sharing ensures latent-space consistency across granularities, while independent encoders preserve view-specific characteristics.

To establish semantic consistency across granularities, we introduce an intra-view contrastive objective. Given two latent representations $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{B\times m}$, we concatenate their row-wise normalized embeddings as $\mathbf{Q} = [\hat{\mathbf{A}}; \hat{\mathbf{B}}]$. The contrastive loss is defined as

$$\ell_{\mathrm{con}}(\mathbf{A}, \mathbf{B}) = -\frac{1}{2B}\sum_{i=1}^{2B} \log \frac{\exp\left(\mathbf{q}_i^\top \mathbf{q}_{p(i)}/\tau\right)}{\sum_{\substack{k=1\\k\neq i}}^{2B} \exp\left(\mathbf{q}_i^\top \mathbf{q}_k/\tau\right)}, \tag{4}$$

where $p(i)$ indexes the positive pair corresponding to the same sample across granularities, and τ is the temperature parameter. Taking the raw branch as the semantic reference, the intra-view objective is

$$\mathcal{L}_{\mathrm{intra}} = \frac{1}{2V}\sum_{v=1}^{V}\left[\ell_{\mathrm{con}}(\mathbf{Z}_v^r, \mathbf{Z}_v^m) + \ell_{\mathrm{con}}(\mathbf{Z}_v^r, \mathbf{Z}_v^g)\right]. \tag{5}$$

This objective aligns sample-level semantics across granularities while retaining their complementary structural information, yielding three coordinated representations for subsequent granularity-aligned cross-view enhancement.

3.2 Granularity-Aligned Enhancement

Given the intra-view representations $\{\mathbf{Z}_v^r, \mathbf{Z}_v^m, \mathbf{Z}_v^g\}$, we perform cross-view interaction only between branches with the same structural granularity. This avoids mixing representations with different semantic roles while allowing each branch to exploit complementary information from other views.

Before information exchange, we encourage same-granularity representations across views to preserve sample-level semantic correspondence:

$$\mathcal{L}_{\mathrm{coh}} = \frac{1}{3\binom{V}{2}}\sum_{b\in\{r,m,g\}}\sum_{1\le u<v\le V} \ell_{\mathrm{con}}\left(\mathbf{Z}_u^b, \mathbf{Z}_v^b\right), \tag{6}$$

where $\ell_{\mathrm{con}}(\cdot,\cdot)$ is defined in Sec. 3.1.

For the b-th granularity of target view v, cross-view information is aggregated from the corresponding branches of the remaining views:

$$\mathbf{C}_v^b = \mathrm{Attn}\left(\mathbf{Z}_v^b, \{\mathbf{Z}_u^b\}_{u\neq v}\right), \qquad b \in \{r, m, g\}. \tag{7}$$

Here, $\mathrm{Attn}(\mathbf{Q}, \{\mathbf{K}_i\})$ denotes scaled dot-product attention in the latent space, with attention weights computed from $\mathbf{Q}\mathbf{K}^\top/\sqrt{d}$, where the candidate representations are stacked into $\mathbf{K}$.

Since the aggregated information may contain irrelevant or redundant components, we introduce a learnable gate to control its contribution:

$$\mathbf{G}_v^b = \sigma\left(\mathcal{G}_v^b\left([\mathbf{Z}_v^b \| \mathbf{C}_v^b]\right)\right), \tag{8}$$

where $\mathcal{G}_v^b(\cdot)$ is a learnable gating mapping, $\mathbf{G}_v^b$ denotes the resulting gate, and $\sigma(\cdot)$ is the sigmoid function. The enhanced representation is

$$\bar{\mathbf{Z}}_v^b = \mathbf{Z}_v^b + \lambda_b \mathbf{G}_v^b \odot \mathbf{C}_v^b, \qquad b \in \{r, m, g\}. \tag{9}$$

Here, λ_b is a branch-specific transfer coefficient, with $\lambda_r = 0.1$, $\lambda_m = 0.3$, and $\lambda_g = 0.2$.

Finally, we use the enhanced raw branch as the semantic anchor and integrate the main and graph branches as complementary structural cues:

$$\mathbf{H}_v = \bar{\mathbf{Z}}_v^r + \lambda_r \operatorname{Attn}\left(\bar{\mathbf{Z}}_v^r, \{\bar{\mathbf{Z}}_v^m, \bar{\mathbf{Z}}_v^g\}\right). \tag{10}$$

Thus, using the same λ_r for raw-centered aggregation, $\mathbf{H}_v$ preserves the raw semantics while incorporating global and local structural information for subsequent low-rank bilinear fusion.

3.3 Low-Rank Bilinear Fusion

After granularity-aligned cross-view enhancement, each view is represented by $\mathbf{H}_v$, which integrates intra-view structural cues and selectively transferred cross-view information. A straightforward fusion strategy is to concatenate these representations and apply a linear transformation. However, purely additive fusion may be insufficient to characterize multiplicative dependencies between representation factors. In high-dimensional spaces, explicitly modeling full bilinear interactions is also computationally expensive, since the pairwise interaction space grows quadratically with the representation dimension.

To balance interaction expressiveness and model complexity, we first construct a residual refinement path for each view and then perform low-rank bilinear fusion between the preserved and refined representations.

Residual bottleneck refinement. Given $\mathbf{H}_v$, we first project it into a compact latent space:

$$\mathbf{T}_v = \operatorname{ReLU}\left(\operatorname{Down}_v(\mathbf{H}_v)\right), \tag{11}$$

where $\operatorname{Down}_v(\cdot)$ denotes a learnable dimensionality reduction mapping.

The compact representation is then projected back and incorporated through a residual connection:

$$\mathbf{R}_v = \mathbf{H}_v + \alpha \operatorname{ReLU}\left(\operatorname{Up}_v(\mathbf{T}_v)\right), \tag{12}$$

where $\operatorname{Up}_v(\cdot)$ denotes the corresponding up-projection and α controls the contribution of the refinement path. This bottleneck-residual transformation refines the learned representation without discarding its original semantic structure, while limiting the additional model capacity.

We then concatenate the enhanced and refined representations across views:

$$\mathbf{H}_s = \operatorname{Concat}\left(\mathbf{H}_1, \ldots, \mathbf{H}_V\right), \qquad \mathbf{R}_s = \operatorname{Concat}\left(\mathbf{R}_1, \ldots, \mathbf{R}_V\right), \tag{13}$$

where $\mathbf{H}_s, \mathbf{R}_s \in \mathbb{R}^{B \times d_s}$ and d_s denotes the concatenated representation dimension.

Low-rank bilinear interaction. For each sample, a full bilinear representation between the corresponding vectors in $\mathbf{H}_s$ and $\mathbf{R}_s$ would involve an outer product in $\mathbb{R}^{d_s \times d_s}$, resulting in d_s^2 pairwise interaction features. Such quadratic expansion is undesirable for high-dimensional representations.

Instead, we project the two representation paths into a low-dimensional interaction space of rank r_f, where $r_f \ll d_s$:

$$\mathbf{M}_1 = \mathbf{U}_f(\mathbf{H}_s), \qquad \mathbf{M}_2 = \mathbf{V}_f(\mathbf{R}_s), \qquad \mathbf{M}_1, \mathbf{M}_2 \in \mathbb{R}^{B \times r_f}. \tag{14}$$

Here, $\mathbf{U}_f(\cdot)$ and $\mathbf{V}_f(\cdot)$ are learnable low-dimensional projections.

The bilinear interaction is factorized as

$$\mathbf{B}_f = \mathbf{M}_1 \odot \mathbf{M}_2. \tag{15}$$

Each dimension of $\mathbf{B}_f$ captures a multiplicative interaction between the corresponding projected factors of the two representation paths. Thus, second-order dependencies can be modeled without explicitly constructing the full outer-product space.

Finally, the interaction features are projected back to the unified representation space and incorporated through a residual connection:

$$\mathbf{F} = \mathbf{H}_s + \beta \mathbf{W}_f \left(\mathbf{B}_f \right), \tag{16}$$

where $\mathbf{W}_f(\cdot)$ maps the low-rank interaction features to the dimension of $\mathbf{H}_s$, and β controls the contribution of the bilinear term.

The residual formulation retains $\mathbf{H}_s$ as the primary information path, while the low-rank bilinear term provides a second-order correction. By replacing the d_s^2 full interaction space with an r_f-dimensional factorized representation, LIVC captures multiplicative dependencies with controlled complexity. The resulting $\mathbf{F}$ serves as the unified representation for clustering.

3.4 Joint Learning Objective and Clustering

LIVC is trained with two complementary objectives: information preservation and semantic coordination. The former is enforced at both the view and fusion levels, while the latter is achieved through the intra-view and cross-view contrastive objectives introduced in Secs. 3.1 and 3.2.

To preserve view-specific information after cross-view enhancement, we reconstruct the standardized raw input $\mathbf{X}_v^r$ from $\mathbf{H}_v$ using a view-specific decoder D_v^H. For two matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{B \times d}$, we define

$$\mathrm{MSE}(\mathbf{A}, \mathbf{B}) = \frac{1}{Bd} \left\| \mathbf{A} - \mathbf{B} \right\|_F^2. \tag{17}$$

The view-level reconstruction loss is

$$\mathcal{L}_{\mathrm{rec}} = \frac{1}{V} \sum_{v=1}^{V} \mathrm{MSE} \left(D_v^H(\mathbf{H}_v), \mathbf{X}_v^r \right). \tag{18}$$

To ensure that the fused representation retains information from all views, we further reconstruct each raw view from $\mathbf{F}$:

$$\mathcal{L}_{\mathrm{fus}} = \frac{1}{V} \sum_{v=1}^{V} \mathrm{MSE} \left(D_v^F(\mathbf{F}), \mathbf{X}_v^r \right), \tag{19}$$

where D_v^F denotes the decoder associated with the v-th view.

The complete objective is

$$\mathcal{L} = \mathcal{L}_{\mathrm{rec}} + \lambda_{\mathrm{fus}} \mathcal{L}_{\mathrm{fus}} + \lambda_{\mathrm{intra}} \mathcal{L}_{\mathrm{intra}} + \lambda_{\mathrm{coh}} \mathcal{L}_{\mathrm{coh}}, \tag{20}$$

where λ_{fus}, λ_{intra}, and λ_{coh} control the contributions of the corresponding terms. The reconstruction losses preserve information at the view and fusion levels, while $\mathcal{L}_{\mathrm{intra}}$ and $\mathcal{L}_{\mathrm{coh}}$ coordinate sample-level semantics across granularities and views, respectively.

After training, the unified representation $\mathbf{F}$ is used for clustering:

$$\hat{\mathbf{y}} = \mathrm{KMeans}(\mathbf{F}, K), \tag{21}$$

where K is the predefined number of clusters. K-means is applied only after representation learning and does not participate in gradient-based optimization. The overall procedure is summarized in Appendix A.2.

4 Experiments

4.1 Experimental Setup

Datasets. We evaluate LIVC on five TCGA multi-omics datasets, including CESC, LGG, UCEC, SARC, and STAD. Each dataset contains four omics views, including copy number alteration (CNA), DNA methylation (MET), mRNA expression, and reverse-phase protein array (RPPA). The multi-omics data used in this study were retrieved in 2024 from the TCGA Genomic Data Commons (GDC) portal (`https://portal.gdc.cancer.gov/`) and the UCSC Xena browser (`https://xenabrowser.net/datapages/`). The five datasets exhibit different sample sizes, numbers of clusters, and feature dimensionalities. The detailed dataset statistics are provided in Appendix A.3.1. The number of samples ranges from 148 to 421, while the dimensionality of several omics views exceeds 20,000, reflecting the typical high-dimensional small-sample characteristics of multi-omics data.Additional implementation details are provided in Appendix A.3.2.

Table 1: Clustering performance comparison (%) on five multi-omics datasets. The best results are highlighted in bold, and the second-best results are underlined.

Method	CESC			LGG			SARC		
	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR
GMAE	56.62 ± 4.95	0.26 ± 0.26	85.81 ± 0.00	41.47 ± 2.14	3.42 ± 2.22	49.98 ± 3.04	33.92 ± 2.43	4.78 ± 2.34	43.61 ± 3.27
MCSF	67.16 ± 1.13	17.25 ± 0.74	85.81 ± 0.00	90.74 ± 0.69	72.50 ± 1.25	90.74 ± 0.69	60.62 ± 3.69	38.41 ± 2.54	63.51 ± 0.92
MISCC	78.38 ± 10.27	15.43 ± 8.34	85.95 ± 0.27	44.23 ± 1.14	1.39 ± 0.19	48.46 ± 0.00	57.01 ± 4.03	28.97 ± 1.97	61.86 ± 1.30
WAH-MVC	59.32 ± 5.71	7.83 ± 7.20	85.81 ± 0.00	88.12 ± 6.02	70.22 ± 10.98	88.31 ± 5.64	58.76 ± 4.47	36.91 ± 2.26	$\mathbf{68.97 \pm 2.20}$
GMVC	63.51 ± 11.62	9.84 ± 5.14	85.81 ± 0.00	42.00 ± 3.58	2.46 ± 0.71	48.55 ± 0.19	55.88 ± 7.31	32.37 ± 9.15	65.15 ± 5.84
SSLNMVC	84.46 ± 3.02	0.65 ± 1.46	85.81 ± 0.00	50.07 ± 3.61	3.16 ± 7.06	50.78 ± 5.21	52.27 ± 9.16	17.24 ± 12.65	53.51 ± 9.40
AEMVC	53.78 ± 2.37	0.79 ± 0.81	85.81 ± 0.00	46.65 ± 4.48	6.98 ± 2.89	53.73 ± 1.05	38.04 ± 3.88	5.63 ± 2.97	45.77 ± 4.84
DMVC-MIC	72.57 ± 1.56	0.15 ± 0.10	85.81 ± 0.00	60.10 ± 0.53	33.60 ± 2.28	60.29 ± 0.11	46.91 ± 1.89	11.42 ± 0.76	53.81 ± 1.84
SCMVC	54.59 ± 7.64	4.71 ± 1.57	85.81 ± 0.00	59.62 ± 15.02	27.45 ± 22.47	62.19 ± 13.68	50.62 ± 5.20	20.84 ± 7.15	58.97 ± 6.00
ACCMVC	60.00 ± 1.75	12.13 ± 2.05	85.81 ± 0.00	62.19 ± 10.92	30.34 ± 19.81	64.37 ± 11.64	43.92 ± 2.37	10.91 ± 2.66	51.34 ± 2.08
SEM	57.30 ± 5.88	6.12 ± 5.72	85.81 ± 0.00	65.23 ± 4.15	42.93 ± 6.99	73.92 ± 4.74	59.07 ± 2.69	39.34 ± 2.91	68.56 ± 2.98
LIVC	$\mathbf{90.27 \pm 3.33}$	$\mathbf{45.99 \pm 7.31}$	$\mathbf{90.27 \pm 3.33}$	$\mathbf{95.15 \pm 0.86}$	$\mathbf{81.37 \pm 2.84}$	$\mathbf{95.15 \pm 0.86}$	$\mathbf{62.68 \pm 2.70}$	$\mathbf{39.57 \pm 1.49}$	66.60 ± 1.10

Method	STAD			UCEC		
	ACC	NMI	PUR	ACC	NMI	PUR
GMAE	28.62 ± 2.04	2.88 ± 0.93	57.23 ± 0.00	30.91 ± 0.88	1.50 ± 0.42	35.90 ± 0.88
MCSF	49.45 ± 0.89	30.75 ± 1.59	68.55 ± 1.27	55.51 ± 4.64	34.17 ± 1.39	61.14 ± 1.51
MISCC	39.81 ± 4.28	2.96 ± 0.62	57.36 ± 0.16	39.60 ± 1.62	6.54 ± 3.54	42.62 ± 4.34
WAH-MVC	47.20 ± 3.25	31.65 ± 2.16	68.68 ± 2.15	54.77 ± 6.80	34.70 ± 4.00	61.88 ± 4.67
GMVC	38.52 ± 6.95	2.36 ± 0.68	57.30 ± 0.13	32.99 ± 2.04	0.95 ± 0.64	33.33 ± 1.99
SSLNMVC	48.75 ± 3.69	10.02 ± 8.99	58.97 ± 1.99	36.69 ± 3.54	3.10 ± 2.16	36.79 ± 3.56
AEMVC	27.33 ± 1.59	3.75 ± 1.50	57.23 ± 0.00	31.65 ± 1.67	1.66 ± 0.58	36.44 ± 1.67
DMVC-MIC	35.88 ± 1.74	16.49 ± 1.52	60.26 ± 1.08	44.00 ± 3.78	15.34 ± 0.93	48.40 ± 3.77
SCMVC	39.87 ± 2.19	21.92 ± 0.74	64.18 ± 0.37	38.81 ± 3.87	10.07 ± 4.70	47.26 ± 4.14
ACCMVC	47.33 ± 8.01	23.47 ± 1.49	64.76 ± 1.10	37.09 ± 2.70	7.64 ± 2.62	42.86 ± 2.94
SEM	44.12 ± 4.63	29.54 ± 3.83	69.90 ± 4.40	46.52 ± 7.62	24.14 ± 8.09	52.84 ± 5.70
LIVC	$\mathbf{52.28 \pm 1.63}$	$\mathbf{39.84 \pm 1.36}$	$\mathbf{75.69 \pm 0.63}$	$\mathbf{60.69 \pm 3.56}$	$\mathbf{39.13 \pm 3.02}$	$\mathbf{65.58 \pm 1.21}$

Comparison Methods. We compare LIVC with eleven recent multi-view clustering methods. GMAE (Zou et al., 2026) disentangles view-common and view-specific representations. MCSF (Cai et al., 2026) integrates spectral learning with multi-level contrastive learning. MISCC (Li et al., 2026b) jointly models intra-view structures and cross-view correlations. WAH-MVC (Wang et al., 2026c) combines hyperbolic representation learning with Wasserstein-based alignment. GMVC (He et al., 2025) integrates adaptive graph learning with graph- and cluster-guided contrastive learning. SSLNMVC (Yan et al., 2025) aligns consensus and view-specific semantics through self-supervised learning. AEMVC (Li et al., 2025) adaptively selects informative feature directions via an activate-then-eliminate mechanism. DMVC-MIC (Cui et al., 2025) employs meta-learned information compression and semantic completion. SCMVC (Wu et al., 2024) performs self-weighted contrastive fusion. ACCMVC (Yan et al., 2024) combines anchor-based representation learning with cluster-level contrastive learning. SEM (Xu et al., 2023) integrates self-weighted contrastive learning with reconstruction regularization.

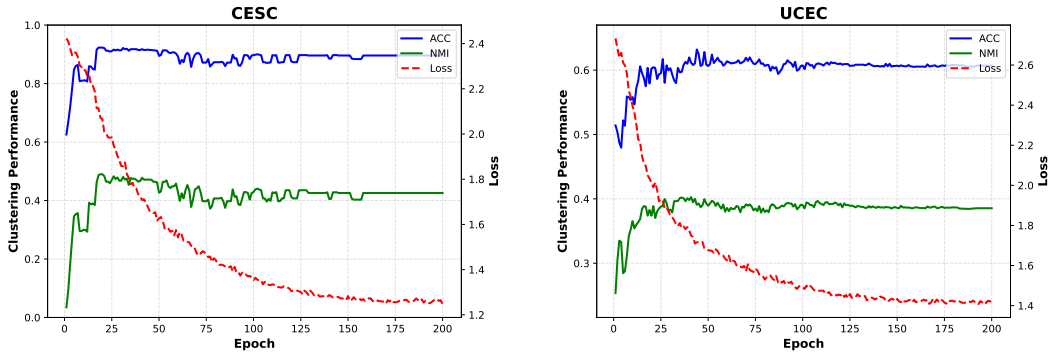
Evaluation Metrics. We evaluate clustering performance using clustering accuracy (ACC), normalized mutual information (NMI), and purity (PUR), which measure the agreement between predicted clusters and ground-truth classes, the mutual dependence between two partitions, and the homogeneity of samples within clusters, respectively. Higher values indicate better clustering performance for all three metrics.

4.2 Overall Clustering Performance

Table 1 reports the clustering performance of LIVC and eleven competing methods on five multi-omics datasets, where all results are presented as the mean $\pm$ standard deviation over five seeds.

LIVC achieves the best mean performance on 14 out of 15 dataset–metric combinations. On the remaining SARC PUR metric, LIVC achieves 66.60%, compared with 68.97% for WAH-MVC and 68.56% for SEM. On CESC, LIVC achieves 90.27% ACC, 45.99% NMI, and 90.27% PUR, exceeding the second-best results by 5.81, 28.74, and 4.32 percentage points, respectively. On LGG, it achieves 95.15% ACC and PUR and 81.37% NMI, outperforming MCSF by 4.41, 4.41, and 8.87 percentage points, respectively.

It is worth noting that several competing methods obtain an identical PUR of 85.81% on CESC, despite showing substantial differences in ACC and NMI. CESC contains 148 samples, with 127



(a) CESC (b) UCEC

Figure 2: Convergence analysis of LIVC on CESC and UCEC. The curves show the evolution of ACC, NMI, and training loss over 200 epochs.

samples in the majority class and 21 in the minority class, giving a majority-class proportion of $127/148 = 85.81\%$. Therefore, when the learned partition is dominated by the majority class, PUR can become saturated at 85.81% even if the underlying clustering structure is not well recovered. In contrast, LIVC achieves a PUR of 90.27% on CESC, exceeding the majority-class proportion of 85.81%, which suggests that its learned partition captures additional class structure beyond a majority-class-dominated assignment.

On SARC, LIVC achieves the highest ACC (62.68%) and NMI (39.57%), while its PUR (66.60%) is lower than those of WAH-MVC (68.97%) and SEM (68.56%). It also achieves the best results across all three metrics on STAD and UCEC, with NMI reaching 39.84% and 39.13%, respectively. These results demonstrate the effectiveness of LIVC in learning discriminative representations across multi-omics datasets with varying sample sizes, cluster structures, and feature dimensionalities.

4.3 MODEL ANALYSIS

Convergence Analysis. As shown in Figure 2, the training loss on both CESC and UCEC decreases rapidly during the early training stage and gradually converges as optimization proceeds. Meanwhile, ACC and NMI increase quickly and remain relatively stable after convergence, indicating that LIVC can effectively optimize the learned representation while maintaining stable clustering performance. Although slight fluctuations can be observed on CESC during the intermediate training stage, the overall optimization process remains stable and eventually reaches a steady state. Results on the remaining datasets are provided in Appendix A.3.3.

Parameter Sensitivity Analysis. We investigate the sensitivity of LIVC to the temperature parameter τ used in the contrastive objectives. As shown in Figure 3, we vary $\tau \in \{0.01, 0.05, 0.1, 0.5, 1, 2\}$ and evaluate the resulting ACC, NMI, and PUR on all five multi-omics datasets. Overall, ACC and PUR remain relatively stable across a broad range of temperature values, whereas NMI exhibits larger variations on several datasets, indicating that semantic consistency is more sensitive to the temperature setting. Nevertheless, LIVC maintains competitive clustering performance across most temperature values, showing a reasonable degree of robustness to τ Eq. 4. Additional sensitivity analyses of λ_{fus}, λ_{intra}, λ_{coh}, α, and β are provided in Appendix A.3.5.

Ablation Study. To investigate the contribution of each major component in LIVC, we conduct component ablation studies on all five datasets, as reported in Table 2. The results in Table 2 are obtained from a single run with one fixed random seed.Specifically, M1 removes all three proposed modules and serves as the basic variant; M2–M4 retain only IVE, GAE, and LRBF, respectively; M5–M7 combine two of the three modules; and M8 denotes the complete LIVC framework with all modules enabled. Overall, the full model achieves the best performance on 11 out of 15 dataset–metric combinations, demonstrating the effectiveness of jointly integrating the three components.

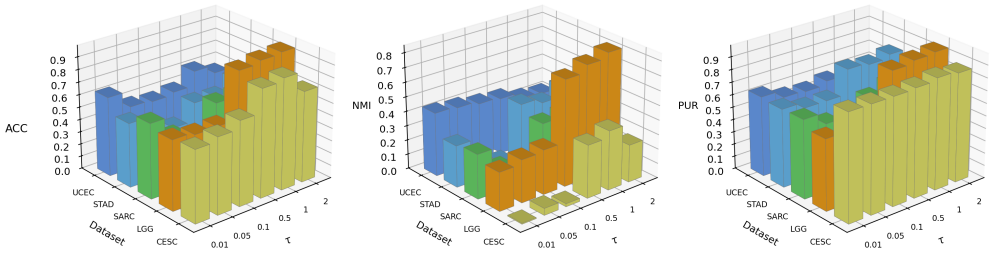


Figure 3: Sensitivity analysis of the temperature parameter τ on five multi-omics datasets in terms of ACC, NMI, and PUR.

Table 2: Ablation study of different components on five multi-omics datasets.

Modules	IVE	GAE	LRBF	CESC			LGG			SARC			STAD			UCEC		
				ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR
M1				65.54	13.30	85.81	93.82	76.73	93.82	62.37	35.22	65.46	38.26	28.39	65.27	55.80	41.70	66.91
M2	✓			87.16	35.79	87.16	94.30	78.91	94.30	63.40	40.93	65.98	42.77	27.89	65.59	57.04	40.33	65.19
M3		✓		76.35	21.21	85.81	95.01	**84.35**	95.01	64.95	38.79	67.53	44.37	25.86	68.49	60.74	40.21	62.72
M4			✓	80.56	28.03	85.81	94.54	79.79	94.54	63.41	36.37	66.49	52.73	35.92	72.03	61.98	37.63	63.46
M5	✓	✓		90.54	43.44	90.54	94.77	79.57	94.77	64.43	42.45	67.53	52.09	40.06	75.88	62.22	36.95	67.16
M6	✓		✓	91.89	50.62	91.89	95.25	81.27	95.25	64.95	**42.97**	67.01	51.44	32.05	67.85	62.72	35.98	64.44
M7		✓	✓	81.76	30.66	85.81	95.49	81.69	95.49	65.46	41.45	**68.56**	54.02	37.86	73.31	63.21	**41.92**	63.46
M8	✓	✓	✓	**94.59**	**57.09**	**94.59**	**95.72**	83.66	**95.72**	**66.49**	41.53	66.49	**54.98**	**41.65**	**76.85**	**65.68**	41.19	**67.65**

Compared with M1, introducing individual modules generally improves clustering performance, although their effects vary across datasets. For example, IVE (M2) substantially improves CESC from 65.54% to 87.16% in ACC and from 13.30% to 35.79% in NMI, indicating the benefit of multi-granularity view expansion. GAE (M3) achieves particularly strong performance on LGG, reaching 95.01% ACC and 84.35% NMI, suggesting that granularity-aligned cross-view interaction can effectively exploit complementary information across views. LRBF (M4) yields clear improvements on the more challenging STAD and UCEC datasets, increasing ACC from 38.26% to 52.73% and from 55.80% to 61.98%, respectively, showing the usefulness of second-order interaction modeling in the fusion stage.

Combining multiple components further improves the overall performance. In particular, M5–M7 generally outperform their corresponding single-module variants on several datasets, indicating that the proposed modules provide complementary benefits. When all three modules are jointly enabled, M8 achieves the highest ACC on all five datasets and the best performance on all three metrics for CESC and STAD. It also obtains the best ACC and PUR on LGG and UCEC. Although several individual metrics favor simpler variants, such as LGG NMI for M3, SARC NMI for M6, SARC PUR for M7, and UCEC NMI for M7, the full model provides the most consistent overall performance across datasets. These results support the complementary roles of IVE, GAE, and LRBF in constructing effective multi-view representations.

5 Conclusion

In this work, we propose LIVC, a multi-view clustering framework designed for high-dimensional small-sample multi-omics data. LIVC first constructs multi-granularity representations from the raw information, global principal structure, and local neighborhood structure through Interactive View Expansion. It then performs cross-view interaction between representations of the same granularity via Granularity-Aligned Enhancement and selectively enhances the raw representation with complementary information. Finally, Low-Rank Bilinear Fusion models second-order interactions between thepreserved and refined representations to obtain a unified representation for clustering.

Future work will investigate the extension of LIVC to incomplete multi-view and missing-omics scenarios, as well as its scalability to larger and more heterogeneous multi-view datasets.

AI USE STATEMENT

In this work, we used generative AI tools only to assist with improving the grammar, readability, clarity, and language consistency of the manuscript. Generative AI tools were used solely for language-level assistance and did not participate in the development of the research content.

We did not use generative AI tools for any research tasks requiring mandatory disclosure. All AI-assisted edits were carefully reviewed, verified, and revised by the authors when necessary.

Generative AI tools were not used to generate research ideas, formulate theoretical claims, design the methodology or experiments, construct proofs, implement the proposed method, or generate, modify, analyze, or interpret experimental results. The methodology, experimental settings, result analysis, and conclusions presented in this work were independently developed by the authors.

We take full responsibility for the final content and accuracy of this work.

ETHICS STATEMENT

This work focuses on methodological research in unsupervised multi-view clustering and does not involve the recruitment of new human subjects, the collection of personally identifiable information, or experimental interventions involving individuals. All experiments are conducted on publicly available benchmark datasets following their established research usage guidelines.

The proposed method is intended for methodological research on representation learning and clustering of high-dimensional multi-view data, and is not directly used for clinical diagnosis, treatment decisions, or individual-level medical assessment. To the best of our knowledge, the proposed method itself does not introduce significant direct ethical risks.

As with other general-purpose representation learning and data-driven analysis methods, potential downstream applications should comply with appropriate requirements regarding data privacy, fairness, transparency, and domain-specific ethical considerations.

REFERENCES

Jintang Bian, Yixiang Lin, Xiaohua Xie, Chang-Dong Wang, Lingxiao Yang, Jian-Huang Lai, and Feiping Nie. Multilevel contrastive multiview clustering with dual self-supervised learning. *IEEE Transactions on Neural Networks and Learning Systems*, 36(6):10422–10436, 2025.

Bing Cai, Xiaoli Wang, Gui-Fu Lu, and Zechao Li. Multi-hierarchical contrastive spectral fusion for multi-view clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 39617–39626, 2026.

Zhe Chen, Xiao-Jun Wu, Tianyang Xu, Hui Li, and Josef Kittler. Multi-layer multi-level comprehensive learning for deep multi-view clustering. *Information Fusion*, 116:102785, 2025. ISSN 1566-2535.

Jinrong Cui, Bang Liufu, Chongjie Dong, Jingcheng Ke, and Jie Wen. Deep multi-view clustering with meta information compression. *IEEE Transactions on Image Processing*, 34:7514–7527, 2025.

Lunke Fei, Junlin He, Qi Zhu, Shuping Zhao, Jie Wen, and Yong Xu. Deep multi-view contrastive clustering via graph structure awareness. *IEEE Transactions on Image Processing*, 34:3805–3816, 2025.

Lele Fu, Sheng Huang, Lei Zhang, Jinghua Yang, Zibin Zheng, Chuanfu Zhang, and Chuan Chen. Subspace-contrastive multi-view clustering. *ACM Trans. Knowl. Discov. Data*, 18(9), October 2024. ISSN 1556-4681.

Jipeng Guo, Xiang Xu, Yu Cao, Man Cao, Mengyuan Xin, Tianxiang Zhao, Ye Su, Junbin Gao, Mingliang Cui, and Youqing Wang. Anchor-to-graph structural co-regularization for scalable multi-view clustering. *Pattern Recognition*, 177:113378, 2026a. ISSN 0031-3203.

Ruiming Guo, Mouxing Yang, Yijie Lin, Xi Peng, and Peng Hu. Robust contrastive multi-view clustering against dual noisy correspondence. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 121401–121421. Curran Associates, Inc., 2024.

Shengzhao Guo, Zhenyu Ma, Jingyu Wang, Feiping Nie, and Xuelong Li. Graph-free multiview clustering with anchors. *IEEE Transactions on Multimedia*, 2026b.

Hongqing He, Jie Xu, Guoqiu Wen, Yazhou Ren, Na Zhao, and Xiaofeng Zhu. Graph embedded contrastive learning for multi-view clustering. In *IJCAI*, pp. 5336–5344, 2025.

Shizhe Hu, Binyan Tian, Weibo Liu, and Yangdong Ye. Self-supervised trusted contrastive multi-view clustering with uncertainty refined. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 17305–17313, 2025.

Zisen Kong, Dongxia Chang, Yiming Wang, Pengyuan Li, and Yao Zhao. Tensorial multi-view clustering via alternative rank minimization and inter-view alignment. *IEEE Transactions on Knowledge and Data Engineering*, 2026.

Guojie Li, Zhiwen Yu, Kaixiang Yang, Jianming Lv, and CL Philip Chen. Pseudo-label similarity graph-driven multi-view contrastive clustering. *IEEE Transactions on Multimedia*, 2026a.

Pengyuan Li, Man Liu, Dongxia Chang, Yiming Wang, Zisen Kong, and Yao Zhao. Aemvc: Mitigate imbalanced embedding space in multi-view clustering. In *Proceedings of the 33rd ACM International Conference on Multimedia*, MM '25, pp. 6461–6470, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400720352.

Pengyuan Li, Dongxia Chang, Yiming Wang, Man Liu, Zisen Kong, Linhua Kong, and Yao Zhao. Deep multi-view clustering with intra-view similarity and cross-view correlation learning. *IEEE Transactions on Multimedia*, 2026b.

Mingyang Liu, Zuyuan Yang, Wei Han, and Shengli Xie. Progressive neighbor-masked contrastive learning for fusion-style deep multi-view clustering. *Neural Networks*, 179:106503, 2024. ISSN 0893-6080.

Yazhou Ren, Zichen Wen, Junlong Ke, Chenhang Cui, Yonghao Huang, Xinyue Chen, Philip S Yu, and Lifang He. Enhancing multi-view clustering: A sufficient information-theoretic approach for consistency acquisition and redundancy elimination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2026.

Chuan Tang, Miaomiao Li, Jun Wang, Chang Tang, En Zhu, and Xinwang Liu. Sample-level weighted and structure-enhanced anchor graph learning for scalable multi-view clustering. *IEEE Transactions on Circuits and Systems for Video Technology*, 2026.

Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

Dexian Wang, Sha Yang, Ping Deng, Zheng Yu, Qiaofeng Wu, Wei Huang, Pengfei Zhang, and Weiping Ding. Deep multi-view clustering algorithm with integrated auto-encoder and non-negative matrix factorization. *Applied Soft Computing*, 190:114552, 2026a. ISSN 1568-4946.

Huibing Wang, Yawei Chen, Mingze Yao, Qian Liu, Guangqi Jiang, and Xianping Fu. Tensorized parameter-free multi-view spectral clustering based on fair representation learning. *IEEE Transactions on Multimedia*, 2026b.

Jing Wang and Songhe Feng. Contrastive and view-interaction structure learning for multi-view clustering. In *IJCAI*, pp. 5055–5063, 2024.

Jing Wang, Songhe Feng, Gengyu Lyu, and Zhibin Gu. Triple-granularity contrastive learning for deep multi-view subspace clustering. In *Proceedings of the 31st ACM International Conference on Multimedia*, MM '23, pp. 2994–3002, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9798400701085.

Rui Wang, Yuting Jiang, Xiaoqing Luo, Xiao-Jun Wu, Nicu Sebe, and Ziheng Chen. Wasserstein-aligned hyperbolic multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 26444–26452, 2026c.

Song Wu, Yan Zheng, Yazhou Ren, Jing He, Xiaorong Pu, Shudong Huang, Zhifeng Hao, and Lifang He. Self-weighted contrastive fusion for deep multi-view clustering. *IEEE Transactions on Multimedia*, 26:9150–9162, 2024.

Yanwanyu Xi, Chang Tang, Jun-Jie Huang, Xingchen Hu, Yuanyuan Liu, and Xinwang Liu. Contrastive and dual adversarial representation learning for multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 2025.

Baili Xiao, Ke Liang, Jiaqi Jin, Jun Wang, Yinbo Xu, Siwei Wang, and En Zhu. Dmcar: Disentangled mixture-of-experts with context-aware routing for multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 27055–27063, 2026.

Huilang Xu, Xiang-Xiang Su, Simin Chen, Guang-Yong Chen, and Xing Chen. Bridging optimization and neural networks for efficient multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 16066–16074, 2026.

Jie Xu, Shuo Chen, Yazhou Ren, Xiaoshuang Shi, Hengtao Shen, Gang Niu, and Xiaofeng Zhu. Self-weighted contrastive learning among multiple views for mitigating representation degeneration. *Advances in neural information processing systems*, 36:1119–1131, 2023.

Weiqing Yan, Yuanyang Zhang, Chang Tang, Wujie Zhou, and Weisi Lin. Anchor-sharing and cluster-wise contrastive network for multiview representation learning. *IEEE Transactions on Neural Networks and Learning Systems*, 36(2):3797–3807, 2024.

Weiqing Yan, Tingyu Yang, and Chang Tang. Self-supervised semantic soft label learning network for deep multi-view clustering. *IEEE Transactions on Multimedia*, 27:4971–4983, 2025.

Zengbiao Yang and Yihua Tan. Deep multi-view clustering based on instance-level adaptive structural contrastive learning. *Neural Networks*, 198:108581, 2026. ISSN 0893-6080.

Yongbo Yu, Zhoumin Lu, Feiping Nie, Weizhong Yu, Zongcheng Miao, and Xuelong Li. Pseudo-label guided bidirectional discriminative deep multi-view subspace clustering. *IEEE Transactions on Knowledge and Data Engineering*, 37(7):4213–4224, 2025.

Hanning Yuan, Zhihui Zhang, Qi Guo, Lianhua Chi, Sijie Ruan, Wei Zhou, Jinhui Pang, and Xiaoshuai Hao. Dwcl: Dual-weighted contrastive learning for robust multi-view clustering. *Engineering Applications of Artificial Intelligence*, 165:113532, 2026. ISSN 0952-1976.

Fenghua Zhang and Hangjun Che. Separable consistency and diversity feature learning for multi-view clustering. *IEEE Signal Processing Letters*, 31:1595–1599, 2024.

Xin Zou, Ruimeng Liu, Chang Tang, Zhenglai Li, Xinwang Liu, Kunlun He, and Wanqing Li. Learning disentangled representations for generalized multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2026.

A Appendix

A.1 Analytical properties of LIVC

This section provides analytical insights into two key components of LIVC. We first characterize the structural selectivity of the granularity-aligned cross-view enhancement mechanism. We then show that the proposed low-rank bilinear fusion implements rank-constrained second-order interactions with reduced computational complexity.

A.1.1 GRANULARITY-ALIGNED SELECTIVE TRANSFER

For granularity $b \in \{r, m, g\}$, LIVC aggregates cross-view information only from branches with the same structural role:

$$\mathbf{C}_v^b = \mathrm{Attn}\left(\mathbf{Z}_v^b, \{\mathbf{Z}_u^b\}_{u \neq v}\right). \tag{22}$$

The corresponding gated residual update is

$$\bar{\mathbf{Z}}_v^b = \mathbf{Z}_v^b + \lambda_b \mathbf{G}_v^b \odot \mathbf{C}_v^b, \tag{23}$$

where

$$\mathbf{G}_v^b = \sigma\left(\mathcal{G}_v^b\left([\mathbf{Z}_v^b \| \mathbf{C}_v^b]\right)\right). \tag{24}$$

Proposition 1 (Granularity-Aligned Selective Transfer)**.** *Assume $\lambda_b \geq 0$. For each target view v and granularity $b \in \{r, m, g\}$, the enhanced representation $\bar{\mathbf{Z}}_v^b$ has direct forward-path dependence only on representations associated with the same granularity b. Moreover,*

$$\left\|\bar{\mathbf{Z}}_v^b - \mathbf{Z}_v^b\right\|_F \leq \lambda_b \left\|\mathbf{C}_v^b\right\|_F. \tag{25}$$

Proof. By construction, $\mathbf{C}_v^b$ is computed only from $\{\mathbf{Z}_u^b\}_{u=1}^V$, and $\mathbf{G}_v^b$ depends only on $\mathbf{Z}_v^b$ and $\mathbf{C}_v^b$. Therefore, within the forward mapping of the enhancement module, $\bar{\mathbf{Z}}_v^b$ does not directly depend on representations from another granularity $b' \neq b$.

Furthermore, the sigmoid activation gives $0 \leq G_{v,ij}^b \leq 1$. Hence,

$$\left|\bar{Z}_{v,ij}^b - Z_{v,ij}^b\right| = \lambda_b \left|G_{v,ij}^b C_{v,ij}^b\right| \leq \lambda_b \left|C_{v,ij}^b\right|. \tag{26}$$

Summing the squared element-wise inequalities yields

$$\left\|\bar{\mathbf{Z}}_v^b - \mathbf{Z}_v^b\right\|_F \leq \lambda_b \left\|\mathbf{C}_v^b\right\|_F. \tag{27}$$

□

Proposition 1 shows that the enhancement module preserves granularity correspondence during direct cross-view information transfer, while the learnable gate bounds the magnitude of the injected residual. This property concerns the forward interaction inside the enhancement module and does not imply independence among different granularities during joint representation learning.

A.1.2 LOW-RANK BILINEAR INTERACTION

For a single sample, let $\mathbf{h}, \mathbf{r} \in \mathbb{R}^{d_s}$ denote the corresponding rows of $\mathbf{H}_s$ and $\mathbf{R}_s$, respectively. Consider the linear components of the low-rank projections:

$$\mathbf{m}_1 = \mathbf{U}\mathbf{h}, \qquad \mathbf{m}_2 = \mathbf{V}\mathbf{r}, \tag{28}$$

where $\mathbf{U}, \mathbf{V} \in \mathbb{R}^{r_f \times d_s}$. The factorized interaction is

$$\mathbf{b} = (\mathbf{U}\mathbf{h}) \odot (\mathbf{V}\mathbf{r}), \tag{29}$$

followed by an output projection $\mathbf{W} \in \mathbb{R}^{d_s \times r_f}$.

Theorem 2 (Rank-Constrained Bilinear Interaction)**.** *For each output dimension j, the factorized interaction $\mathbf{W}\mathbf{b}$ can be written as*

$$[\mathbf{W}\mathbf{b}]_j = \mathbf{h}^\top \mathbf{A}_j \mathbf{r}, \tag{30}$$

where

$$\mathbf{A}_j = \sum_{k=1}^{r_f} W_{jk} \mathbf{u}_k \mathbf{v}_k^\top, \tag{31}$$

with $\mathbf{u}_k^\top$ and $\mathbf{v}_k^\top$ denoting the k-th rows of $\mathbf{U}$ and $\mathbf{V}$, respectively. Furthermore,

$$\mathrm{rank}(\mathbf{A}_j) \leq r_f. \tag{32}$$

Proof. The k-th element of $\mathbf{b}$ is

$$b_k = (\mathbf{u}_k^\top \mathbf{h})(\mathbf{v}_k^\top \mathbf{r}). \tag{33}$$

Therefore,

$$[\mathbf{W}\mathbf{b}]_j = \sum_{k=1}^{r_f} W_{jk}(\mathbf{u}_k^\top \mathbf{h})(\mathbf{v}_k^\top \mathbf{r}) \tag{34}$$

$$= \mathbf{h}^\top \left(\sum_{k=1}^{r_f} W_{jk} \mathbf{u}_k \mathbf{v}_k^\top \right) \mathbf{r} \tag{35}$$

$$= \mathbf{h}^\top \mathbf{A}_j \mathbf{r}. \tag{36}$$

Since each matrix $\mathbf{u}_k \mathbf{v}_k^\top$ has rank at most one,

$$\mathrm{rank}(\mathbf{A}_j) \le r_f. \tag{37}$$

□

Theorem 2 shows that the proposed factorization retains explicit second-order multiplicative interactions while restricting each output-specific interaction matrix to rank at most r_f.

For a mini-batch of size B, computing the two low-dimensional projections and the output projection requires

$$O(Bd_s r_f) \tag{38}$$

operations, with

$$O(d_s r_f) \tag{39}$$

learnable parameters, ignoring biases. In contrast, explicitly constructing the full outer-product interaction requires $O(Bd_s^2)$ interaction features. Therefore, when $r_f \ll d_s$, the proposed factorization provides second-order modeling without explicitly forming the quadratic interaction space.

Finally, LIVC uses the residual formulation

$$\mathbf{F} = \mathbf{H}_s + \beta \mathbf{W}_f(\mathbf{B}_f), \tag{40}$$

so that the original first-order representation is preserved while the low-rank bilinear term acts as a second-order correction.

A.2 OPTIMIZATION PROCEDURE

The optimization procedure is summarized in Algorithm 1.

A.3 ADDITIONAL EXPERIMENTAL RESULTS

A.3.1 DATASETS DESCRIPTION

Our experiments are conducted on five TCGA multi-omics datasets, including CESC, LGG, UCEC, SARC, and STAD. All datasets contain four omics views, i.e., copy number alteration (CNA), DNA methylation (MET), mRNA expression, and reverse-phase protein array (RPPA). The five datasets differ in sample size, number of clusters, and feature dimensionality, while all exhibit typical high-dimensional small-sample characteristics. Detailed information of these datasets is summarized in Table 3.

A.3.2 IMPLEMENTATION DETAILS

Experiments are conducted on an Intel Core i7-12700K CPU, 32 GB RAM, and an NVIDIA GeForce RTX 4070 GPU (12 GB), using Python 3.8.20, PyTorch 2.3.1, and CUDA 11.8. For each view, the raw, main, and graph branches share a view-specific encoder with a hidden-layer width of 128 and ReLU activation to map different granularities into a unified latent space. LIVC is optimized using Adam with an initial learning rate of 5×10^{-4} and a batch size of 128. A cosine annealing learning-rate scheduler is adopted, and the model is trained for 200 epochs. The weights

Algorithm 1 Training and Clustering Procedure of LIVC

Require: Multi-view data $\{\mathbf{X}_v\}_{v=1}^{V}$, number of clusters K, training epochs E, and model hyper-parameters.
Ensure: Unified representation $\mathbf{F}$ and clustering assignments $\hat{\mathbf{y}}$.

Intra-view structural construction: for each view v, construct the raw, main, and graph branches $\mathbf{X}_v^r$, $\mathbf{X}_v^m$, and $\mathbf{X}_v^g$.
Initialize all learnable parameters.
for $e = 1, \ldots, E$ **do**
  for each mini-batch **do**
    Encode the three intra-view granularities using the shared view-specific encoder: $\mathbf{Z}_v^b = E_v(\mathbf{X}_v^b)$, $b \in \{r, m, g\}$.
    Compute the intra-view cross-granularity objective $\mathcal{L}_{\text{intra}}$.
    Compute the granularity-aligned cross-view coherence objective $\mathcal{L}_{\text{coh}}$.
    Perform granularity-aligned cross-view aggregation and selective transfer to obtain $\{\bar{\mathbf{Z}}_v^r, \bar{\mathbf{Z}}_v^m, \bar{\mathbf{Z}}_v^g\}_{v=1}^{V}$.
    Refine the raw semantic anchor with the enhanced main and graph branches to obtain $\{\mathbf{H}_v\}_{v=1}^{V}$.
    Apply residual bottleneck refinement to obtain $\{\mathbf{R}_v\}_{v=1}^{V}$.
    Concatenate the multi-view representations: $\mathbf{H}_s = \mathrm{Concat}(\mathbf{H}_1, \ldots, \mathbf{H}_V)$ and $\mathbf{R}_s = \mathrm{Concat}(\mathbf{R}_1, \ldots, \mathbf{R}_V)$.
    Compute low-rank bilinear interaction $\mathbf{B}_f = \mathbf{U}_f(\mathbf{H}_s) \odot \mathbf{V}_f(\mathbf{R}_s)$, and obtain $\mathbf{F} = \mathbf{H}_s + \beta \mathbf{W}_f(\mathbf{B}_f)$.
    Compute the view-level and fusion-level reconstruction objectives $\mathcal{L}_{\text{rec}}$ and $\mathcal{L}_{\text{fus}}$.
    Compute $\mathcal{L} = \mathcal{L}_{\text{rec}} + \lambda_{\text{fus}}\mathcal{L}_{\text{fus}} + \lambda_{\text{intra}}\mathcal{L}_{\text{intra}} + \lambda_{\text{coh}}\mathcal{L}_{\text{coh}}$.
    Update all learnable parameters using Adam.
  end for
end for
Extract the unified representation $\mathbf{F}$ for all samples.
$\hat{\mathbf{y}} \leftarrow \mathrm{KMeans}(\mathbf{F}, K)$.
return $\mathbf{F}$ and $\hat{\mathbf{y}}$.

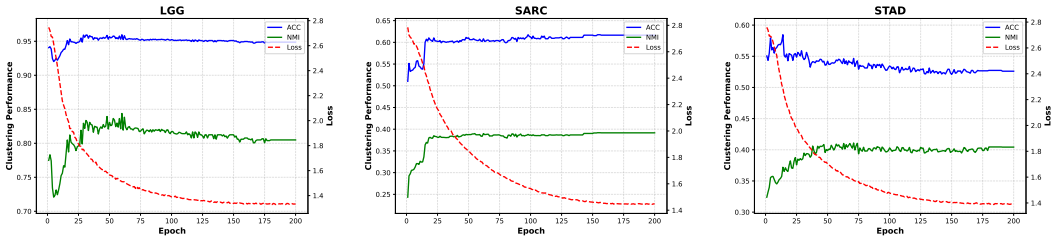
Table 3: Statistics of the multi-omics datasets used in experiments.

Dataset	Samples	Clusters	Views	Dimensionality
CESC	148	2	4	24157/22564/20020/193
LGG	421	3	4	19452/22588/20163/190
UCEC	405	4	4	17912/22598/17507/191
SARC	194	4	4	24751/22538/20154/193
STAD	311	5	4	24480/22582/16765/193

of the reconstruction loss, fusion reconstruction loss, intra-view contrastive loss, and cross-view coherence loss are set to 1.0, 1.0, 0.1, and 0.05, respectively.

The cumulative explained-variance threshold for PCA reconstruction is set to $\rho = 0.95$, and KNN smoothing uses $k = 10$ neighbors. The latent representation dimension is set to 64, the bottleneck dimension for residual refinement is 16, and the low-rank interaction dimension is set to $r_f = 64$. The temperature parameter in the contrastive objectives is fixed to $\tau = 1.0$. The branch-specific cross-view transfer coefficients are set to $\lambda_r = 0.1$, $\lambda_m = 0.3$, and $\lambda_g = 0.2$ for the raw, main, and graph branches, respectively. The residual refinement coefficient and low-rank bilinear fusion coefficient are set to $\alpha = 0.1$ and $\beta = 0.1$, respectively.

After the final training epoch, the trained model is used to re-extract the unified representation $\mathbf{F}$ of all samples, which is subsequently fed into K-means for clustering. No ground-truth-based clustering metric is used for epoch or checkpoint selection. Unless otherwise stated, the main comparison experiments are repeated with five random seeds, and the mean and standard deviation of ACC, NMI, and PUR are reported. The component ablation study in Table 2 uses one fixed random seed.



(a) LGG (b) SARC (c) STAD

Figure 4: Additional convergence analysis of LIVC on LGG, SARC, and STAD. The curves show the evolution of ACC, NMI, and training loss over 200 epochs.

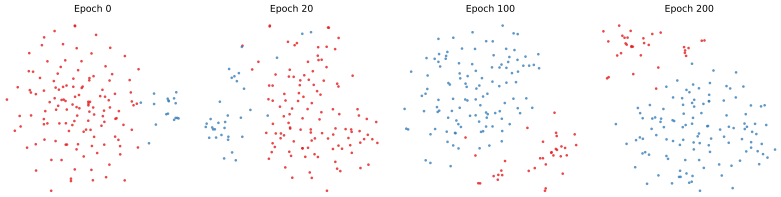


Figure 5: t-SNE visualization of the unified representation $\mathbf{F}$ learned by LIVC on CESC at Epoch 0, 20, 100, and 200. Different colors denote different ground-truth classes.

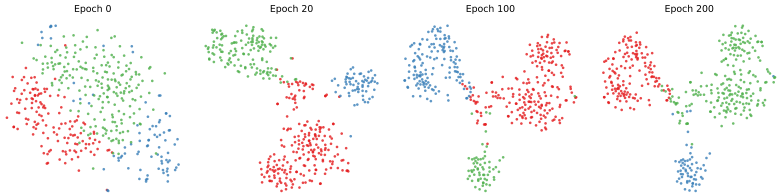


Figure 6: t-SNE visualization of the unified representation $\mathbf{F}$ learned by LIVC on LGG at Epoch 0, 20, 100, and 200. Different colors denote different ground-truth classes.

A.3.3 ADDITIONAL CONVERGENCE ANALYSIS

To further examine the optimization behavior of LIVC, we provide additional convergence results on LGG, SARC, and STAD. As shown in Figure 4, the training loss on all three datasets decreases progressively and gradually approaches a stable value as optimization proceeds. Meanwhile, ACC and NMI improve rapidly during the early training stage and become relatively stable in later epochs. Although slight fluctuations can be observed on some datasets, the overall optimization process remains stable. Together with the convergence results on CESC and UCEC reported in the main text, these results indicate that LIVC exhibits consistent convergence behavior across datasets with different clustering characteristics.

A.3.4 VISUALIZATION ANALYSIS OF FEATURE CLUSTERING PROCESS

We employ t-SNE (Van der Maaten & Hinton, 2008) to visualize the unified representation $\mathbf{F}$ learned by LIVC at different training epochs on all five multi-omics datasets. Figures 5–9 show the representation distributions at Epoch 0, 20, 100, and 200. At the beginning of training, samples from different classes are relatively scattered and exhibit substantial overlap. As optimization proceeds, samples from the same class gradually form more coherent groups, while the separation between different classes becomes increasingly evident. Although some overlap remains on the more challenging datasets, the overall cluster structures become progressively clearer. These results provide an intuitive illustration of how LIVC gradually learns clustering-friendly representations throughout the training process.

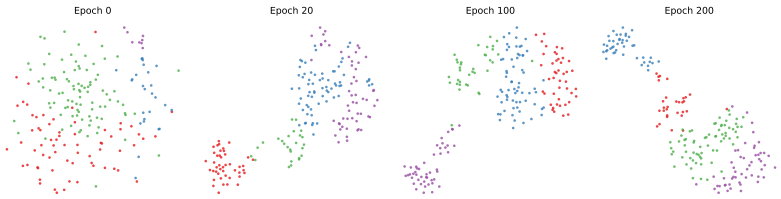


Figure 7: t-SNE visualization of the unified representation $\mathbf{F}$ learned by LIVC on SARC at Epoch 0, 20, 100, and 200. Different colors denote different ground-truth classes.

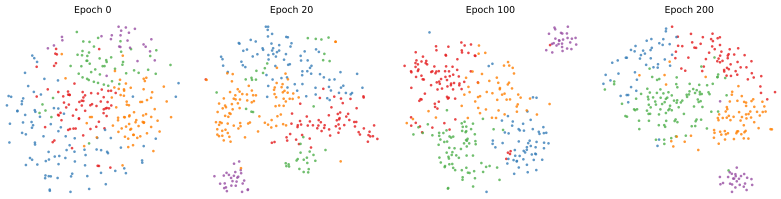


Figure 8: t-SNE visualization of the unified representation $\mathbf{F}$ learned by LIVC on STAD at Epoch 0, 20, 100, and 200. Different colors denote different ground-truth classes.

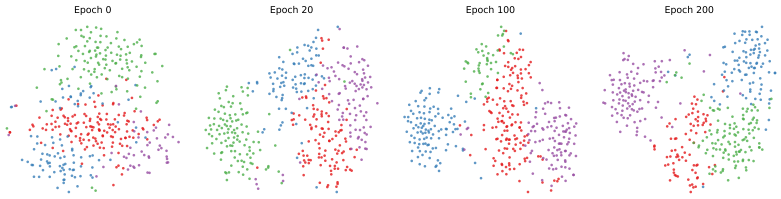


Figure 9: t-SNE visualization of the unified representation $\mathbf{F}$ learned by LIVC on UCEC at Epoch 0, 20, 100, and 200. Different colors denote different ground-truth classes.

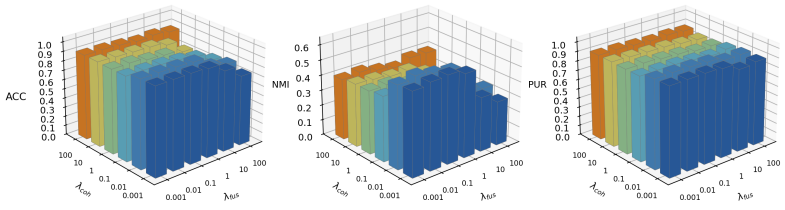
A.3.5 ADDITIONAL PARAMETER SENSITIVITY ANALYSIS

To further investigate the robustness of LIVC to different hyperparameter settings, we conduct additional joint parameter sensitivity analyses on all five multi-omics datasets. Specifically, we consider four parameter pairs: $(\lambda_{\text{fus}}, \lambda_{\text{coh}})$, $(\lambda_{\text{fus}}, \lambda_{\text{intra}})$, $(\lambda_{\text{intra}}, \lambda_{\text{coh}})$, and (α, β). Here, λ_{fus}, λ_{intra}, and λ_{coh} correspond to the loss weights defined in Eqs. 19, 5, and 6, respectively, while α and β control the residual bottleneck refinement and low-rank bilinear interaction in Eqs. 12 and 16. For each parameter pair, the remaining hyperparameters are fixed to their default values, while ACC, NMI, and PUR are used to evaluate the resulting clustering performance.

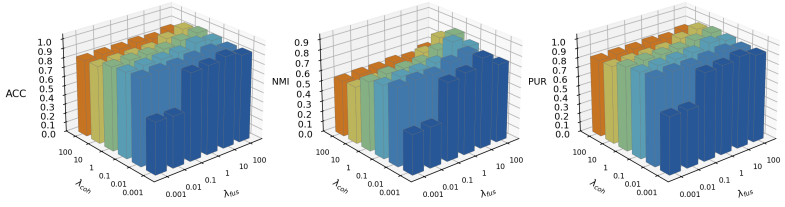
Figures 10–13 present the corresponding results on CESC, LGG, SARC, STAD, and UCEC. Overall, the effect of different parameter combinations varies across datasets and evaluation metrics. ACC and PUR generally remain relatively stable over a broad range of settings, whereas NMI shows larger fluctuations under some configurations. This suggests that the clustering structure measured by NMI is relatively more sensitive to the balance among different objectives and enhancement coefficients.

For the loss-weight combinations, the results indicate that neither excessively weak nor excessively strong weighting consistently benefits all datasets. Instead, moderate parameter settings generally provide a better balance between reconstruction, intra-view consistency, and cross-view coherence. Similarly, the joint analysis of α and β shows that the strengths of residual bottleneck refinement and low-rank bilinear interaction should be properly balanced, since their relative contributions vary across datasets.

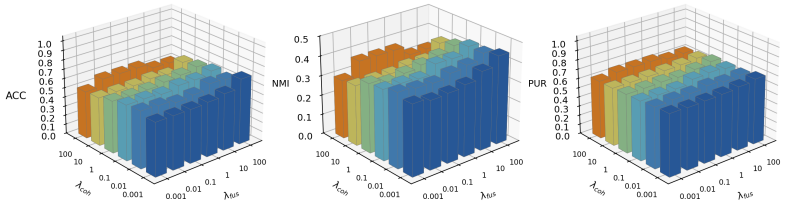
Although the optimal parameter combination is not identical for every dataset and metric, LIVC maintains competitive clustering performance across a broad range of configurations. Together with the temperature sensitivity analysis in the main text, these results demonstrate that LIVC does not



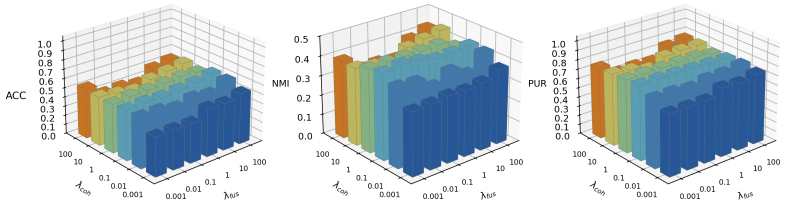
(a) CESC



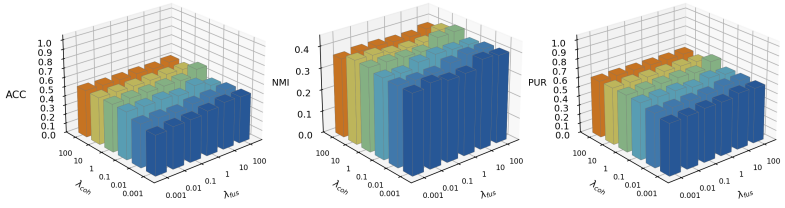
(b) LGG



(c) SARC



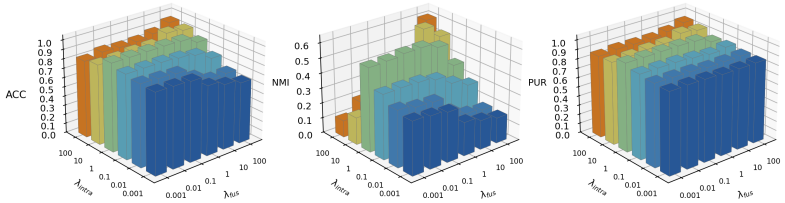
(d) STAD



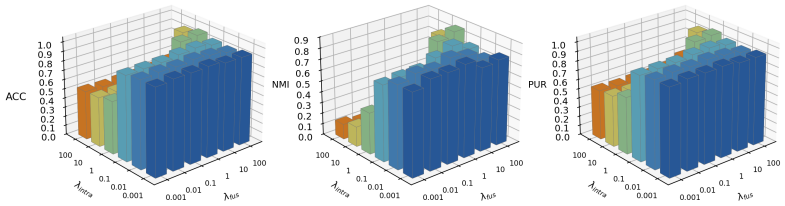
(e) UCEC

Figure 10: Sensitivity analysis of λ_{fus} and λ_{coh} on five multi-omics datasets.

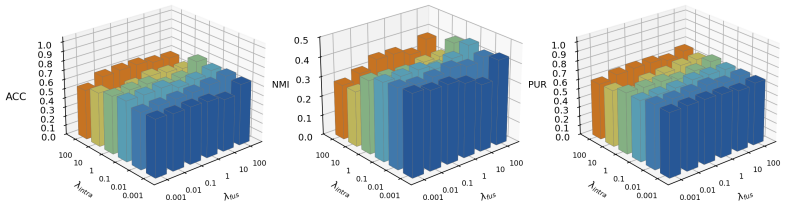
rely on a narrowly tuned hyperparameter setting while also highlighting the importance of reasonable parameter selection.



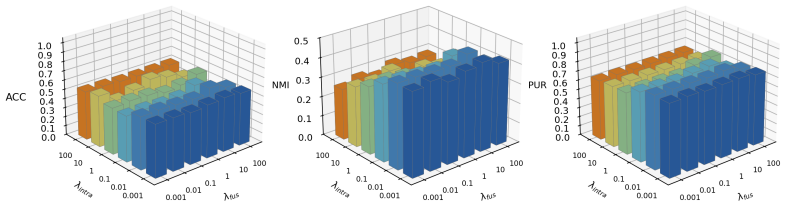
(a) CESC



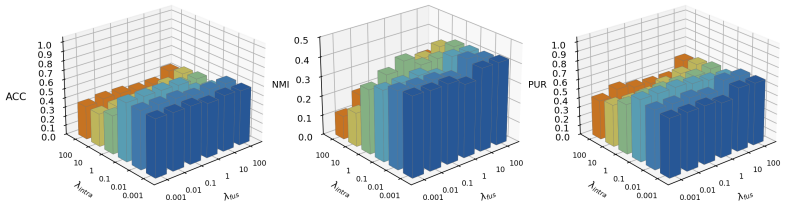
(b) LGG



(c) SARC

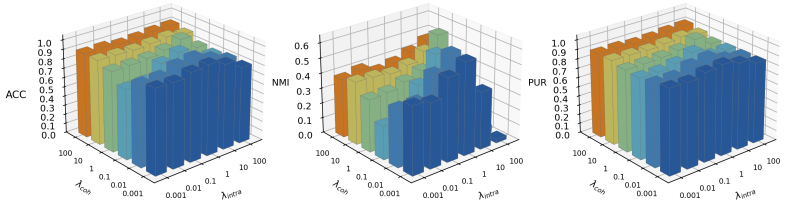


(d) STAD

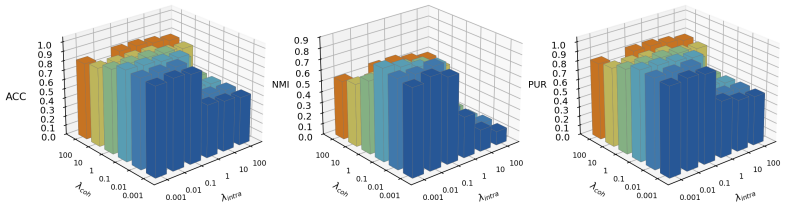


(e) UCEC

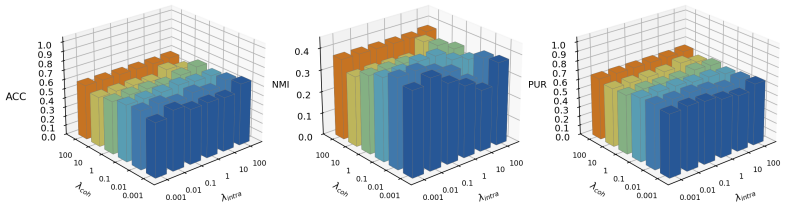
Figure 11: Sensitivity analysis of λ_{fus} and λ_{intra} on five multi-omics datasets.



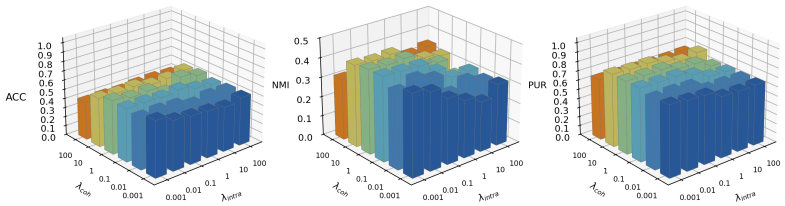
(a) CESC



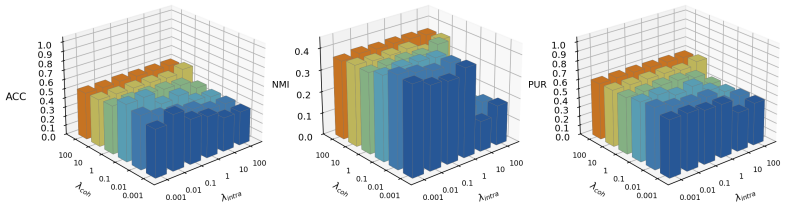
(b) LGG



(c) SARC

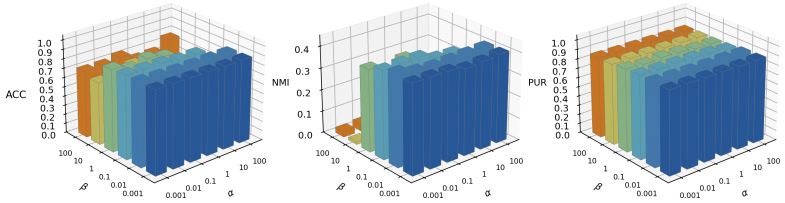


(d) STAD

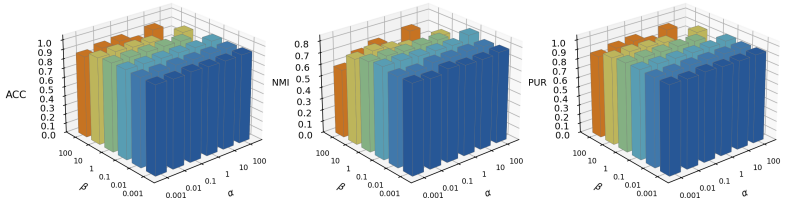


(e) UCEC

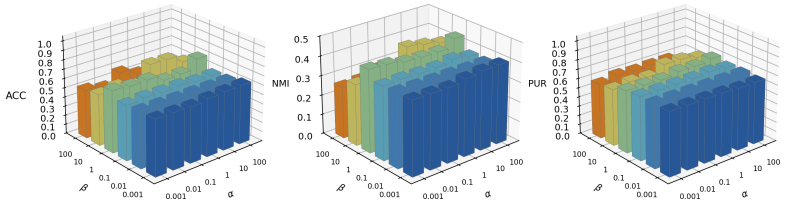
Figure 12: Sensitivity analysis of λ_{intra} and λ_{coh} on five multi-omics datasets.



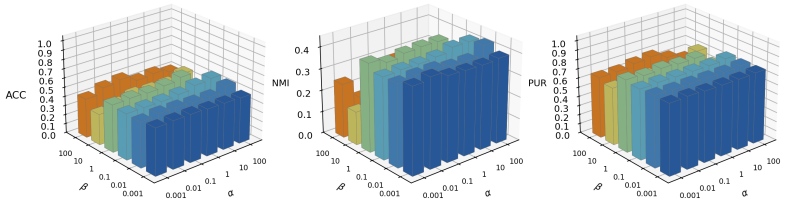
(a) CESC



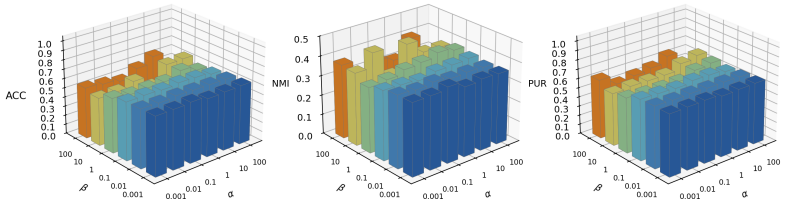
(b) LGG



(c) SARC



(d) STAD



(e) UCEC

Figure 13: Sensitivity analysis of α and β on five multi-omics datasets.