

Prototype Instance-semantic Disentanglement with Low-rank Regularized Subspace Clustering for WSIs Explainable Recognition

Chentao Li¹ and Pan Huang²(✉)

¹ Fu Foundation School of Engineering, Columbia University, New York, USA

² Centre for Smart Health, School of Nursing, The Hong Kong Polytechnic University, Hong Kong SAR, China

✉ Corresponding author: Pan Huang (panhuang@polyu.edu.hk)

Abstract. The tumor region plays a key role in pathological diagnosis. Tumor tissues are highly similar to precancerous lesions and non-tumor instances often greatly exceed tumor instances in whole-slide images (WSIs). These issues cause instance-semantic entanglement in multi-instance learning frameworks, degrading both model representation capability and interpretability. To address this, we propose an end-to-end prototype instance-semantic disentanglement framework with low-rank regularized subspace clustering, *i.e.* PID-LRSC, in two aspects. First, we use secondary instance-subspace learning to construct low-rank regularized subspace clustering (LRSC), addressing instance entanglement caused by an excessive proportion of non-tumor instances. Second, we employ enhanced contrastive learning to design prototype instance-semantic disentanglement (PID), resolving semantic entanglement caused by the high similarity between tumor and precancerous tissues. We conduct extensive experiments on multicentre pathology datasets, implying that PID-LRSC outperforms other state-of-the-art methods. Overall, PID-LRSC provides clearer instance semantics during decision-making and significantly enhances the reliability of auxiliary diagnostic outcomes. The code and datasets are available at <https://github.com/Prince-Lee-PathAI/PID-LRSC>.

Keywords: Prototype learning · Whole-slide image · Pathological grading · Multiple Instance Learning · Subspace Clustering.

1 Introduction

Tumor regions serve as the primary basis for pathological grading, as clinical decisions are largely determined within malignant tissue [23]. However, in whole-slide images (WSIs), tumor areas account for only a very small proportion and they are highly similar to precancerous lesions in terms of morphological characteristics. This instance-semantic entanglement has intrinsically resulted in weak representation capability and limited interpretability of existing multi-instance learning (MIL) frameworks.

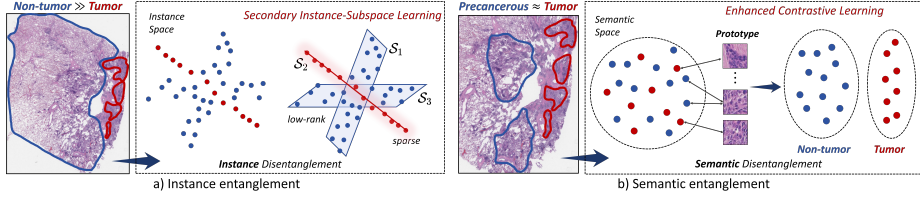


Fig. 1. The motivation of proposed PID-LRSC. **Left:** Instance entanglement caused by excessive proportion of non-tumor over tumor. **Right:** Semantic entanglement caused by high similarity between precancerous lesions and tumor.

Instance entanglement arises from the extreme imbalance within WSIs, where non-tumor regions overwhelmingly dominate but truly discriminative tumor instances are sparsely distributed [9]. In Figure 1, in the instance space, feature representations of non-tumor regions are highly redundant across different subspaces and thus noise accurate WSI representation. Without distinguishing and suppressing such irrelevant regions, the model fails to focus more on tumor-related subspaces and reduce instance-level subspace entanglement. For example, ACMIL [26] introduces multi-branch attention and stochastic top-k masking to capture diverse informative instances. DGR-MIL[28] employed a diverse global representation strategy to MIL framework by modeling the diversity among instances through a set of global vectors. These models [7,15,10,3,13,25,21,16,1] ignore sparse tumor regions and are incapable of addressing instance entanglement. Meanwhile, biased MIL aggregation process will also overestimate redundant non-tumor instances and inhibit sparse tumor instances by assigning approximately equal weights, which weakens the model’s representational capacity and pathological grading performance.

Additionally, semantic entanglement stems from the high morphological similarity between precancerous lesions and malignant tumor cells, as illustrated in Figure 1. In the semantic space, the ambiguity of tumor-related semantics significantly limits the interpretability of MIL models. Lack of external pathological criteria, the models cannot correctly identify the semantic category to which each instance belongs, failing to perform accurate posterior semantic attribution. Consequently, instance-level explanations become unreliable and may highlight irrelevant low-effect regions instead of truly discriminative tumor areas. For example, MiCo[12] can enhance cross-regional intra-tissue correlations and strengthen inter-tissue semantic associations in WSIs with context-aware clustering. FuzzyMIL [14] proposed a deep fuzzy clustering framework based on learnable fuzzy C-means to analyze WSIs in a compact and efficient manner. These cluster-based MIL models [20,8,4,24,27,19] have attempted to address instance entanglement by distinguishing subspaces, but they merely approximate instance pseudo labels and suffer from semantically entangled feature representation, degrading both interpretability and classification performance.

Incorporating external knowledge, prototype-driven learning can effectively alleviate poor model interpretability. PHIM-MIL [22] introduce prototypes for similar weight aggregation and multi-scale feature fusion. Gou et.al [6] proposed

a vision-language based queryable prototype MIL framework for incremental WSI classification. Moreover, HCF-MIL [11] is a prototype-driven MIL model with clustering to filter irrelevant non-tumor noise. These models either ignore the sparsity of tumor semantics for effective subspace clustering, or simply calculating prototype similarities, not exploring instance-semantic disentanglement.

For this purpose, we introduce PID-LRSC, an end-to-end instance-semantic disentangled learning framework for explainable WSIs recognition. First, we develop a low-rank regularized subspace clustering (LRSC) method for secondary instance-subspace exploration. Two coarse-to-fine low-rank constraints inhibit non-tumor features and enables sparse tumor features to form discriminative subspaces, alleviating entanglement caused by imbalanced instance distribution. Second, we implement prototype instance-semantic disentanglement (PID) with enhanced contrastive learning. By comparing with prototypes, the model tends to focus more on high-effect areas and learns semantically disentangled decision boundaries between tumor and precancerous lesions. Through end-to-end training, our PID-LRSC not only disentangles instances into domain-specific semantics, but also improves the classification performance and interpretability.

2 Methodology

2.1 Problem Statement

Assume that differentiation states are hypothetically divided into tumor (TIs), non-tumor (NTIs) and background instances (BGIs) in a WSI, where tumor instances are in small proportion, yet are more correlated with pathological grading than non-tumor instances.

Instead of patch label fusion, We adopt instance aggregation based MIL framework that aims to learn a permutation-invariant scoring function $\mathcal{S}(\cdot)$ that maps the set of instances $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ to label $\mathbf{Y}$:

$$\hat{\mathbf{Y}} = \mathcal{S}(\mathbf{X}) \quad \text{where} \quad \mathcal{S}(\mathbf{X}) = f(\mathcal{A}(\mathcal{F}(\mathbf{X}))) \quad (1)$$

where $\mathcal{A}(\cdot)$, $f(\cdot)$ denote the aggregation function, prediction head with softmax normalization, respectively. $\mathcal{F}(\cdot)$ denotes the feature extractor.

2.2 Two-phase Coarse-to-fine LRSC

Sparse tumor regions are within large proportion of non-tumor regions in WSIs. Inspired by deep clustering and metric learning, we proposed a two-phase coarse-to-fine adaptive clustering approach with low-rank constraint to obtain robust subspace results. The general form of metric learning is as followed:

$$d_{\mathbf{A}}(\mathbf{z}_1, \mathbf{z}_2) = \sqrt{(\mathbf{z}_1 - \mathbf{z}_2)^T \mathbf{A} (\mathbf{z}_1 - \mathbf{z}_2)} \quad (2)$$

where $\mathbf{A} \in \mathbb{R}^{n \times n}$ is a trainable positive semi-definite matrix and $\mathbf{z}_1, \mathbf{z}_2 \in \mathbb{R}^{n \times 1}$ denote the instance feature representation vectors.

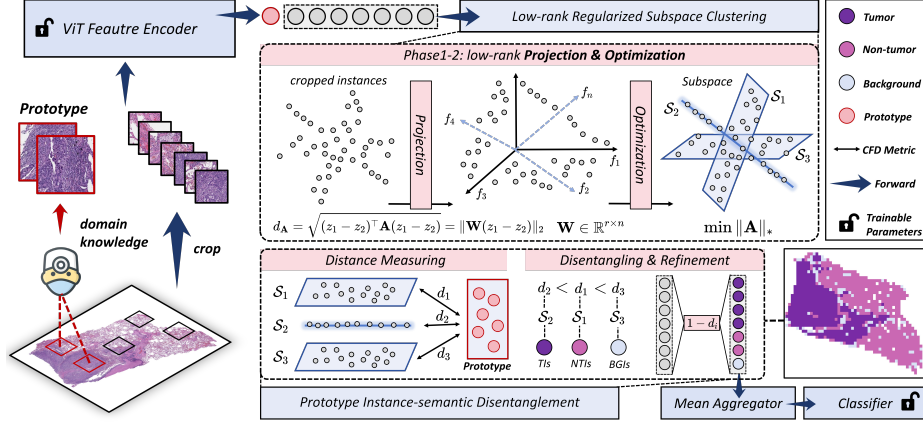


Fig. 2. Overview of **PID-LRSC**. Cropped instances extracted by encoder are clustered into three subspaces by **Low Rank Subspace Clustering** method in two phases. They are then identified into tumor, non-tumor and background by **Prototype Instance-semantic Disentanglement**. End-to-end optimization is performed for disentangled representation learning with refined instance features.

Phase 1: Low-rank Projection. Positive semi-definite matrix ensures the symmetry and non-negativity of the metric. It contributes to robust distance measuring in high dimensional space. Thus, we replace the L_p norm with the adaptive metric $d(\cdot, \cdot)$ and ensure positive semi-definite $\mathbf{A}$ by parameterizing $\mathbf{A} = \mathbf{W}^\top \mathbf{W}$, $\mathbf{W} \in \mathbb{R}^{r \times n}$. Furthermore, we set $r \ll n$ yielding a training-efficient metric. It also serves as a low-rank projection that searches for the best distribution-aware subspaces to obtain robust clustering results; see Fig. 2.

$$d_{\mathbf{A}}(z_1, z_2) = \sqrt{(z_1 - z_2)^\top \mathbf{W}^\top \mathbf{W} (z_1 - z_2)} = \|\mathbf{W}(z_1 - z_2)\|_2 \quad (3)$$

where $\text{rank}(\mathbf{A}) \leq \min(\text{rank}(\mathbf{W}), \text{rank}(\mathbf{W}^\top)) \leq r$. So $\mathbf{W}$ puts a hard low-rank constraint on $\mathbf{A}$ that directly and coarsely embeds features in a lower dimensional subspaces with less redundancy.

Phase 2: Low-rank Optimization. In Phase 1, we manually set a fixed r , but excessively small r will restrict model's capability of representation. Since rank function is non-convex, we apply nuclear norm regularization as a soft constraint to guarantee rank minimization; see Fig. 2 and the following derivation.

$$\begin{aligned} \|\mathbf{A}\|_* &= \sum_{i=1}^n \sigma_i(\mathbf{A}) = \sum_{i=1}^n \sqrt{\lambda_i(\mathbf{A}^\top \mathbf{A})} = \sum_{i=1}^n \sqrt{\lambda_i(\mathbf{A}^2)} = \sum_{i=1}^n \sqrt{(\lambda_i(\mathbf{A}))^2} \\ &= \sum_{i=1}^n |\lambda_i(\mathbf{A})| = \sum_{i=1}^n \lambda_i(\mathbf{A}) = \text{Tr}(\mathbf{A}) \end{aligned} \quad (4)$$

where $\|\cdot\|_*$ denotes nuclear norm, and $\sigma_i(\cdot)$, $\lambda_i(\cdot)$ denote singular values, eigenvalues respectively. By minimizing this regularization term in the loss function, it fine-tunes the model to adaptively approximate the optimal rank associated with varying data distribution.

Overall, this coarse-to-fine LRSC incorporates the optimal $\mathbf{A}^* = \mathbf{W}^{*\top} \mathbf{W}^*$ into KMeans, improving instance entanglement for TIs, NTIs, and BGIs.

2.3 Latent Characteristic function discrepancy based PID

After the LRSC clustering process, we obtain three subspaces, i.e. $\{\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3\} \leftarrow \mathbf{Z} = \mathcal{F}(\mathbf{X})$. We also denote prototype instance subspace $\mathbf{Z}^{\text{PIs}}$, which serves as the anchor to address instance-semantic entanglement. Conventional clustering-based MIL models perform mean-pooling or fixed weights for aggregation, which neglects sample diversities and instances distinctions. Consequently, a distribution-aware metric can cover more feature diversities for substantial samples.

Distance Measuring. Characteristic function discrepancy (CFD) is for robust assessment of distributional discrepancies in dataset distillation. It decouples both phase and amplitude information to ensure dataset realism and diversity [18]. Unlike maximum mean discrepancy (MMD), CFD enables flexible and latent kernel selection by solving a min-max problem:

$$d_k = D(\mathbf{Z}_k, \mathbf{Z}^{\text{PIs}}) = \mathbb{E}_{\mathbf{z} \sim \mathbf{Z}_k, \mathbf{z}' \sim \mathbf{Z}^{\text{PIs}}} \int_{\mathbf{t}} \sqrt{\text{Chf}(\mathbf{t})} dF_{\mathcal{T}}(\mathbf{t}; \psi) \quad (5)$$

$$\text{Chf}(\mathbf{t}) = (|\Phi_{\mathbf{z}}(\mathbf{t}) - \Phi_{\mathbf{z}'}(\mathbf{t})|^2 + (2|\Phi_{\mathbf{z}}(\mathbf{t})||\Phi_{\mathbf{z}'}(\mathbf{t})| \times (1 - \cos(\mathbf{a}_{\mathbf{z}}(\mathbf{t}) - \mathbf{a}_{\mathbf{z}'}(\mathbf{t}))))$$

where $F_{\mathcal{T}}(\mathbf{t}; \psi)$ is the sampling distribution on $\mathbf{t}$ with sampling function ψ and $\mathbf{a}_{\mathbf{z}}(\mathbf{t})$, $\Phi_{\mathbf{z}}(\mathbf{t})$ are phase, amplitude function, respectively. Here $k = 1, 2, 3$.

Given feature representations $z \in \mathbb{R}^{768}$ and randomly sampled frequency vectors $t \sim F_{\mathcal{T}}(t; \psi)$, the characteristic function is formulated as:

$$\varphi_z(t) = \mathbb{E}_z [e^{it^\top z}] = \mathbb{E}_z [\cos(t^\top z) + i \sin(t^\top z)] \quad (6)$$

Accordingly, the real and imaginary components are defined as $\Re(\varphi_z(t)) = \mathbb{E}_z [\cos(t^\top z)]$ and $\Im(\varphi_z(t)) = \mathbb{E}_z [\sin(t^\top z)]$. Based on the complex representation, the amplitude function is computed as:

$$\Phi_z(t) = |\varphi_z(t)| = \sqrt{\Re(\varphi_z(t))^2 + \Im(\varphi_z(t))^2}, \quad (7)$$

while the phase function is obtained by:

$$a_z(t) = \arctan \left(\frac{\Im(\varphi_z(t))}{\Re(\varphi_z(t))} \right). \quad (8)$$

In practice, the expectation over characteristic frequencies is approximated via Monte Carlo sampling using 4096 randomly sampled frequency vectors.

Disentangling and Refinement. We can disentangle the three clustered subspaces based on the CFD metric above and refine the WSI representation. The

smallest distance d_k means stronger effect of $\mathbf{Z}_k$ for SCC grading, i.e. tumor areas. Conversely, it is identified as background for the largest distance. Finally, we get the map with labels {TIs, NTIs, BGIs} and employ the normalized distances to refine the corresponding instance features to get the ultimate representation $\mathbf{Z}_{\text{WSI}}$.

$$\mathbf{Z}_{\text{WSI}} = \left(\mathbf{Z}^{\text{TIs}}, \mathbf{Z}^{\text{NTIs}}, \mathbf{Z}^{\text{BGIs}}, \mathbf{Z}^{\text{PIs}} \right) \cdot \begin{pmatrix} 1 - d_{\min} \\ 1 - d_{\text{med}} \\ 1 - d_{\max} \\ 1 \end{pmatrix} \quad (9)$$

where $d_{\min}, d_{\text{med}}, d_{\max}$ are the minimum, median, maximum value of d_k .

2.4 End-to-end Optimization

End-to-end optimization can learn more domain-specific feature representations. Specifically, we formulate the loss function as follows:

$$\begin{cases} \mathcal{L} = \mathcal{L}_{ce}(f(\mathcal{A}(\mathbf{Z}_{\text{WSI}}))) + \gamma_1 \cdot c_{\min} - \gamma_2 \cdot c_{\max} + \gamma_3 \cdot \text{Tr}(\mathbf{A}) \\ c_{\min} = d_{\mathbf{A}}(\text{mean}(\mathbf{Z}^{\text{TIs}}), \text{mean}(\mathbf{Z}^{\text{PIs}})) \\ c_{\max} = d_{\mathbf{A}}(\text{mean}(\mathbf{Z}^{\text{BGIs}}), \text{mean}(\mathbf{Z}^{\text{PIs}})) \end{cases} \quad (10)$$

where $\mathcal{L}_{ce}$ denotes the cross entropy. Since KMeans clustering is not normally differentiable, the cluster-separation regularizer $c_{\min}$ and $c_{\max}$ encourage tumor-instance cluster approach our prototype instances, optimizing the trainable matrix $\mathbf{A}$ by computing the gradients $\frac{\partial \mathcal{L}}{\partial \mathbf{W}}$.

3 Experimental Results

3.1 Datasets and Setup

To validate PG-CIDL for diagnosis and sub-typing tasks, we conduct extensive experiments on two public datasets (DHMC-KIDNEY, DHMC-LUNG) and two private datasets from Army Medical University. **AMU-LSCC** is for laryngeal squamous cell carcinoma pathological grading and **AMU-CSCC** is for cervical squamous cell carcinoma. **AMU-LSCC** dataset includes 342 whole slide images that was divided by a 6:4 training-validation ratio for each category. Specially Grade I contains a total of 89 WSIs, Grade II contains 152 WSIs and Grade III includes 101 WSIs. **AMU-CSCC** dataset includes 262 whole slide images, where Grade I contains a total of 27 WSIs, Grade II contains 127 WSIs and Grade III includes 108 WSIs.

Prototype annotations were provided by five experts under a double-blind setting using the union set of annotated tumor regions. 64 prototypes per WSI are included in AMU-CSCC, 81 for AMU-LSCC and 5 for the rest public datasets. The prototype instances are class-specific only for training stage. Since the annotations are constructed from multi-expert unions, they inherently include annotation variability. Sufficient quantity of prototype from different level of pathologists also alleviates sensitivity to prototype quality.

Table 1. Performance comparison of cancer diagnosis, and sub-typing on AMU-CSCC and AMU-LSCC with Swin-T pretrained features. DHMC-KIDNEY and DHMC-LUNG are for generalization.

Models	AMU-LSCC		AMU-CSCC		DHMC-KIDNEY		DHMC-LUNG	
	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC
ABMIL [7]	0.6739	0.7888	0.8302	0.9547	0.7675	0.9421	0.7333	<u>0.8123</u>
CLAM-SB [15]	0.6957	0.8367	0.8302	0.9581	0.7807	0.9339	0.7556	0.7783
CLAM-MB [15]	0.6812	0.8031	0.8774	0.9336	0.7939	0.9372	0.7333	0.7684
TransMIL [16]	0.8261	0.9021	0.8868	0.9521	0.8465	0.9603	0.7333	0.8084
DTFD-MIL [25]	0.5435	0.7245	0.8113	0.9036	0.7061	0.9159	0.7333	0.8084
IBMIL-DS [13]	0.6377	0.7672	0.7830	0.8881	0.7939	0.9145	0.6889	0.7091
ILRA-MIL [21]	0.6884	0.8048	0.8491	0.9123	0.8333	0.8693	0.7778	0.8074
S4MIL [5]	0.7971	0.8845	0.8491	0.9605	0.8158	0.9188	0.7111	0.7007
FRMIL [1]	0.6377	0.8020	0.8113	0.8990	0.7807	0.9174	0.7111	0.7442
DGR-MIL [28]	0.8188	0.9147	0.8868	0.9613	<u>0.8596</u>	<u>0.9680</u>	<u>0.8000</u>	0.7886
ACMIL-MHA [26]	<u>0.8551</u>	<u>0.9219</u>	<u>0.9057</u>	0.9701	0.8553	0.9560	0.6889	0.6849
RRT-MIL [17]	0.7681	0.8732	0.8585	0.9265	0.8421	0.9378	0.7778	0.7521
MFC-MIL [2]	0.8261	0.9126	0.8962	<u>0.9723</u>	0.8116	0.9456	0.7778	0.7767
PID-LRSC (ours)	0.9130	0.9590	0.9245	0.9891	0.8991	0.9821	0.8222	0.8711

Table 2. Ablation experiments for different components of PID-LRSC on AMU-LSCC.

Models	Cluster	LRSC	PID	ACC	AUC
ACMIL-MHA	–	–	–	0.8551	0.9219
	✗	✗	✗	<u>0.8696</u>	0.9567
PID-LRSC	✓	✗	✗	0.8116	0.9455
	✓	✓	✗	0.8623	<u>0.9581</u>
	✓	✓	✓	0.9130	0.9590

3.2 Comparison with SOTA Methods

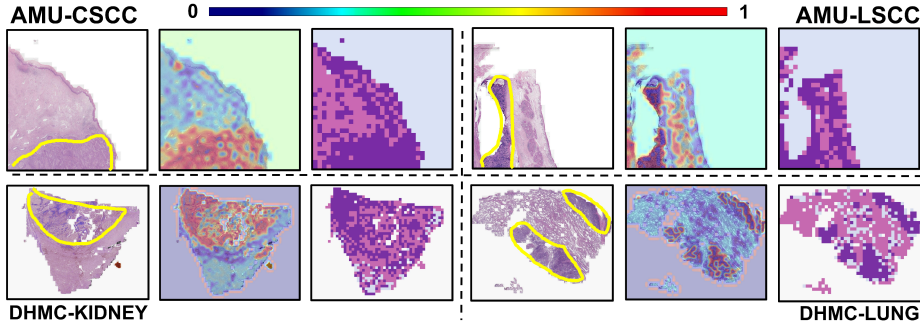
Table 1 illustrates the sub-typing results of MIL approaches on two private datasets and two public datasets, which suggests that our proposed PID-LRSC has superior pathological grading performance. ACMIL-MHA achieves sub-optimal performance on AMU-LSCC, while PID-LRSC improves the ACC and AUC by 5.79% and 3.71%, respectively. For AMU-CSCC, the improvement is 1.88% and 1.68%. To further validate the capability of generalization, we compare those models on multicentre datasets. DGR-MIL achieves the second-best ACC and AUC on DMHC-KIDNEY, while our PID-LRSC improves them by 3.95% and 1.41%, respectively. DHMC-LUNG has seen similar trend with 2.22% ACC improvement on AMU-CSCC and 5.88% AUC improvement.

3.3 Ablation Experiments

In Table 2, without any component as Cluster, LRSC or PID, our baseline model achieves higher ACC and AUC compared to second-best ACMIL-MHA on AMU-

Table 3. Ablation experiments on different distribution metrics in PID.

Metrics in PID (AMU-LSCC)	ACC	AUC
MMD	0.8551	0.9441
CFD (ours)	0.9130	0.9590

**Fig. 3.** Visualization results on multicentre datasets. The first column is original WSI with pathologist-annotated boundaries. The second is the attention contribution scores corresponding to its category. The last column is the disentangling results.

LSCC. It indicates that end-to-end training with acceptable resources can learn better task-specific knowledge than two-stage framework. Moreover, directly incorporated L2-norm clustering in high dimensional space and fixed weights, the AUC performance only reaches 0.9455, 1.12% drop compared to baseline. When LRSC is introduced, the model performance improves significantly, with ACC increasing from 0.8116 to 0.8623 and AUC from 0.9455 to 0.9581. However, without the PID module, the clustering labels remain ambiguous. Therefore, by combining both the LRSC and PID modules, our model achieves the highest ACC 0.9130 and AUC 0.9590. Moreover, Table 3 shows the optimal distribution alignment of CFD compared to MMD.

3.4 Visualization of WSI Interpretability and Representation

We provide visualization results of our PID-LRSC on multicentre datasets in Figure 3. We can tell that both disentangling results and attention heatmaps are highly aligned with pathologists' annotations.

An effective deep classification model is capable of capturing differences in feature distributions, which can be visualized through the analysis of variance (ANOVA) plots. Therefore, We employ UMAP to visualize the final layer feature representation of models in 1-D space. In ANOVA, a larger effect size η^2 indicates a stronger capacity of the model to capture and distinguish class-specific differences. Figure 4 demonstrates that PID-LRSC shows distinct distributions boundaries for SCC grading tasks, while ACMIL-MHA overlaps within different classes. Statistically, it has larger effect size 0.5637 and 0.7884 on both SCC datasets. Despite that DGR-MIL has larger η value, our model still reaches a

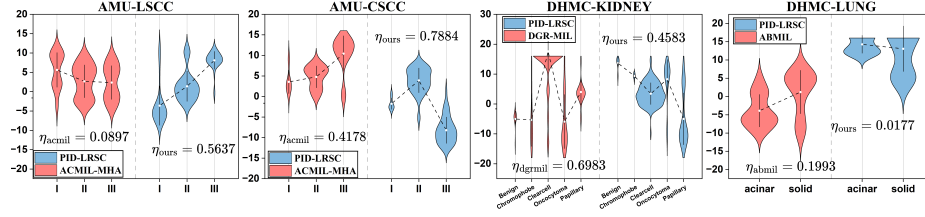


Fig. 4. ANOVA plots on four datasets where η_* denote the value of effect size. Statistically, $\eta_* > 0.4$ implies the model having large effects within categories.

considerable level of effect. Therefore, we can conclude that our PID-LRSC has strong capability of feature representation.

4 Conclusion

Tumor regions play a pivotal role in pathological diagnosis, yet the high visual similarity between tumor and precancerous tissues, together with the overwhelming proportion of non-tumor instances in WSIs, leads to severe instance-semantic entanglement in MIL frameworks. Our proposed PID-LRSC, integrating low-rank regularized subspace clustering and prototype instance-semantic disentanglement effectively disentangles semantic ambiguity between tumor and precancerous tissues. Extensive experiments on multicentre pathology datasets demonstrate that PID-LRSC consistently outperforms state-of-the-art methods, achieving superior classification performance and interpretability. Overall, PID-LRSC provides more explicit instance semantics and robust decision evidence, substantially enhancing the reliability and clinical value of WSI-based diagnosis.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Chikontwe, P., Kim, M., Jeong, J., Sung, H.J., Go, H., Nam, S.J., Park, S.H.: Fr-mil: Distribution re-calibration based multiple instance learning with transformer for whole slide image classification. *IEEE Trans. Med. Imaging* (2024)
2. Cui, X., Chen, W., Su, J.: A multiscale frequency domain causal framework for enhanced pathological analysis. In: *The Thirteenth International Conference on Learning Representations* (2025)
3. Das, K., Conjeti, S., Roy, A.G., Chatterjee, J., Sheet, D.: Multiple instance learning of deep convolutional neural networks for breast histopathology whole slide classification. In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. pp. 578–581. *IEEE* (2018)
4. Du, G., Zhou, L., Li, Z., Wang, L., Lü, K.: Neighbor-aware deep multi-view clustering via graph convolutional network. *Inf. Fusion* **93**, 330–343 (2023)

5. Fillioux, L., Boyd, J., Vakalopoulou, M., Cournède, P.H., Christodoulidis, S.: Structured state space models for multiple instance learning in digital pathology. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 594–604. Springer (2023)
6. Gou, J., Ji, L., Liu, P., Ye, M.: Queryable prototype multiple instance learning with vision-language models for incremental whole slide image classification. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 39, pp. 3158–3166 (2025)
7. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International Conference on Machine Learning. pp. 2127–2136. PMLR (2018)
8. Ju, W., Gu, Y., Chen, B., Sun, G., Qin, Y., Liu, X., Luo, X., Zhang, M.: Glcc: A general framework for graph-level clustering. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 4391–4399 (2023)
9. Kumar, N., Gupta, R., Gupta, S.: Whole slide imaging (WSI) in pathology: Current perspectives and future directions. *J. Digit. Imaging* **33**(4), 1034–1040 (2020)
10. Li, B., Li, Y., Eliceiri, K.W.: Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14318–14328 (2021)
11. Li, C., Huang, P., Qin, J., Luo, X.: Knowledge-driven multiple instance learning with hierarchical cluster-incorporated aware filtering for larynx pathological grading. *IEEE Journal of Biomedical and Health Informatics* **30**(4), 2973–2985 (2026)
12. Li, J., Liu, J., Kuang, H., Yue, H., He, M., Wang, J.: Mico: Multiple instance learning with context-aware clustering for whole slide image analysis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 376–385. Springer (2025)
13. Lin, T., Yu, Z., Hu, H., Xu, Y., Chen, C.W.: Interventional bag multi-instance learning on whole-slide pathological images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 19830–19839 (2023)
14. Liu, A., Li, T., Cai, J., Vajrала, S.S.V.: Fuzzymil: Decoupling pathological phenotypes through deep fuzzy clustering for efficient whole slide image analysis. In: ICASSP 2025–2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1–5. IEEE (2025)
15. 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. *Nat. Biomed. Eng.* **5**(6), 555–570 (2021)
16. Shao, Z., Bian, H., Chen, Y., Wang, Y., Zhang, J., Ji, X., Zhang, Y.: Transmil: Transformer based correlated multiple instance learning for whole slide image classification. *Advances in Neural Information Processing Systems* **34**, 2136–2147 (2021)
17. Tang, W., Zhou, F., Huang, S., Zhu, X., Zhang, Y., Liu, B.: Feature re-embedding: Towards foundation model-level performance in computational pathology. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11343–11352 (2024)
18. Wang, S., Yang, Y., Liu, Z., Sun, C., Hu, X., He, C., Zhang, L.: Dataset distillation with neural characteristic function: A minmax perspective. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 25570–25580 (2025)
19. Wang, X., Cai, D., Yang, S., Cui, Y., Zhu, J., Wang, K., Zhao, J.: Sac-net: enhancing spatiotemporal aggregation in cervical histological image classification via label-efficient weakly supervised learning. *IEEE Trans. Circuits Syst. Video Technol.* (2023)

20. Wu, W., Gao, C., DiPalma, J., Vosoughi, S., Hassanpour, S.: Improving representation learning for histopathologic images with cluster constraints. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 21404–21414 (2023)
21. 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)
22. Xie, Y., Liu, Z., Zhao, J., Ma, J.: Phim-mil: Multiple instance learning with prototype similarity-guided feature fusion and hard instance mining for whole slide image classification. *Information Fusion* **117**, 102847 (2025)
23. Yan, G., Li, L., Zhu, S., Wu, Y., Liu, Y., Zhu, L., Zhao, Z., Wu, F., Jia, N., Liao, C., Jiang, L., Zeng, Q., Wang, P., Shi, L., Zheng, Z., Fang, S., Zhang, G., Tang, Y., Wang, X.: Single-cell transcriptomic analysis reveals the critical molecular pattern of UV-induced cutaneous squamous cell carcinoma. *Cell Death Dis.* **13**(1), 23 (2021)
24. Yang, X., Liu, Y., Zhou, S., Wang, S., Tu, W., Zheng, Q., Liu, X., Fang, L., Zhu, E.: Cluster-guided contrastive graph clustering network. In: Proceedings of the AAAI conference on Artificial Intelligence. vol. 37, pp. 10834–10842 (2023)
25. Zhang, H., Meng, Y., Zhao, Y., Qiao, Y., Yang, X., Coupland, S.E., Zheng, Y.: Dtdf-mil: Double-tier feature distillation multiple instance learning for histopathology whole slide image classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18802–18812 (2022)
26. Zhang, Y., Li, H., Sun, Y., Zheng, S., Zhu, C., Yang, L.: Attention-challenging multiple instance learning for whole slide image classification. In: European Conference on Computer Vision. pp. 125–143. Springer (2024)
27. Zheng, Y., Wu, K., Li, J., Tang, K., Shi, J., Wu, H., Jiang, Z., Wang, W.: Partial-label contrastive representation learning for fine-grained biomarkers prediction from histopathology whole slide images. *IEEE J. Biomed. Health Inform.* (2024)
28. Zhu, W., Chen, X., Qiu, P., Sotiras, A., Razi, A., Wang, Y.: Dgr-mil: Exploring diverse global representation in multiple instance learning for whole slide image classification. In: European Conference on Computer Vision. pp. 333–351. Springer (2024)