MiCo: Multiple Instance Learning with Context-Aware Clustering for Whole Slide Image Analysis

Junjian Li[1], Jin Liu[1,2*], Hulin Kuang[1], Hailin Yue[1], Mengshen He[1], and Jianxin Wang[1,2]

[1] Hunan Provincial Key Lab on Bioinformatics, School of Computer Science and Engineering, Central South University, Changsha 410083, Hunan, China
[2] Xinjiang Engineering Research Center of Big Data and Intelligent Software, School of Software, Xinjiang University, Urumqi 830091, China
liujin06@csu.edu.cn

Abstract. Multiple instance learning (MIL) has shown significant promise in histopathology whole slide image (WSI) analysis for cancer diagnosis and prognosis. However, the inherent spatial heterogeneity of WSIs presents critical challenges, as morphologically similar tissue types are often dispersed across distant anatomical regions. Conventional MIL methods struggle to model these scattered tissue distributions and capture cross-regional spatial interactions effectively. To address these limitations, we propose a novel Multiple instance learning framework with Context-Aware Clustering (MiCo), designed to enhance cross-regional intra-tissue correlations and strengthen inter-tissue semantic associations in WSIs. MiCo begins by clustering instances to distill discriminative morphological patterns, with cluster centroids serving as semantic anchors. To enhance cross-regional intra-tissue correlations, MiCo employs a Cluster Route module, which dynamically links instances of the same tissue type across distant regions via feature similarity. These semantic anchors act as contextual hubs, propagating semantic relationships to refine instance-level representations. To eliminate semantic fragmentation and strengthen inter-tissue semantic associations, MiCo integrates a Cluster Reducer module, which consolidates redundant anchors while enhancing information exchange between distinct semantic groups. Extensive experiments on two challenging tasks across nine large-scale public cancer datasets demonstrate the effectiveness of MiCo, showcasing its superiority over state-of-the-art methods. The code is available at https://github.com/junjianli106/MiCo.

Keywords: Multiple Instance Learning · Whole Slide Image · Context-Aware Clustering.

* means the corresponding author.

1 Introduction

Whole Slide Images (WSIs) provide comprehensive and detailed representations of cellular morphology and the tumor microenvironment, both of which are crucial for accurate diagnostic evaluation and prognostic assessment [2,3,21,6]. However, histopathological WSIs often exhibit extremely high resolutions, reaching up to 100,000 $\times$ 100,000 pixels. The high resolution of WSIs, along with the absence of pixel-level annotations, presents significant challenges in their analysis and modeling [12,14,9,17,8]. Although Multiple Instance Learning (MIL) has emerged as a paradigm for weakly supervised WSI analysis [1,20], existing methods still struggle to handle the spatial complexity of WSIs.

Despite their capability in extracting discriminative features from local image patches, current foundation models still struggle with the challenge of spatial heterogeneity in pathological structures. This core limitation manifests at two interrelated levels: a failure to capture both fine-grained cellular arrangements and broader tissue architecture. At the macroscopic level, key structures of the same histological type are often discontinuously distributed as multi-focal sites across distant regions. For instance, lesions of lymphovascular invasion or perineural invasion may span multiple anatomical planes [11,13]. This spatial separation hinders the model from learning the intrinsic correlations among instances that belong to the same pathological class but are located in different regions. At the microscopic level, each of these distributed structures interacts with a highly heterogeneous cellular microenvironment, forming complex local pathological patterns that require simultaneous modeling of local discriminative features and cross-regional semantic dependencies.

Despite significant progress, existing MIL methods still struggle to model long-range dependencies between pathologically related regions that are morphologically distant, a challenge arising from the inherent spatial heterogeneity of WSIs. For instance, traditional attention-based MIL methods (e.g., AMIL [5]) leverage attention mechanisms to dynamically identify and aggregate critical instances, which tend to focus on instance-level feature learning but fail to capture the semantic relationships between distant pathological regions. Transformer-based methods (e.g., TransMIL [15]) use transformers to learn the global contextual information of WSIs but treat all interactions as homogeneous, failing to capture dynamic heterogeneity interactions across the tumor microenvironment. Graph-based methods (e.g., PatchGCN [1]) rely on fixed neighborhood definitions, while hierarchical methods like HVTSurv [16] restrict local learning to pre-determined sub-regions, both inadequate for modeling tissue-level morphological continuities across spatial distributions.

To overcome these limitations, we propose MiCo, a novel Multiple Instance Learning framework with Context-Aware Clustering. MiCo employs a Cluster Route module to enhance intra-tissue semantic associations by aggregating and propagating information from dispersed patches of the same tissue type, refining instance-level representations. Additionally, MiCo integrates a Cluster Reducer module to eliminate semantic fragmentation and strengthen inter-tissue semantic associations, consolidating semantically redundant anchors while preserving

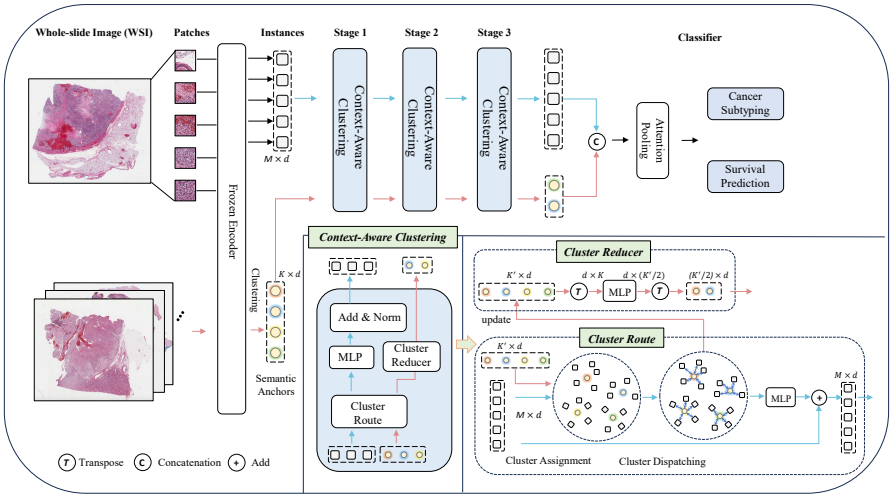

Fig. 1: Overview of MiCo. MiCo consists of multi-layered context-aware clustering modules. Each module is organized by a Cluster Route module, which aggregates and propagates semantic information, and a Cluster Reducer module, which consolidates redundant anchors while enhancing information exchange between semantically distinct anchors.

pathological diversity. Extensive experiments on two challenging tasks across nine large-scale public cancer datasets validate MiCo's effectiveness, showcasing its superior performance over state-of-the-art methods in survival prediction and cancer subtyping tasks.

2 Methodology

2.1 Overview of MiCo

To address the challenges of spatial heterogeneity in histopathological WSIs, we propose MiCo, a novel MIL framework with Context-Aware Clustering. As illustrated in Fig.1, MiCo consists of multi-layered context-aware clustering modules, each organized by two core components: the Cluster Route (CluRoute) module and the Cluster Reducer (CluReducer) module. MiCo begins by clustering instances to distill discriminative morphological patterns, with cluster centroids serving as semantic anchors. The CluRoute dynamically links instances of the same tissue type across distant regions based on feature similarity, aggregating and propagating semantic information to refine instance-level features. The CluReducer consolidates redundant anchors while facilitating information exchange between distinct semantic groups, thereby eliminating semantic fragmentation and strengthening inter-tissue semantic associations. Then, the refined instances and anchor features are fused, and an attention pooling mechanism is applied to enable the model to focus on the most task-relevant information.

2.2 WSI Preprocessing

Given a WSI $\boldsymbol{W}_j$, we first crop the tissue regions into non-overlapping patches $\{p_{j,m}\}_{m=1}^{M}$, where the total number of patches varies per WSI. Each patch is encoded into a feature vector $h_{j,m} \in \mathbb{R}^d$ using the CONCH v1.5 [4], a multimodal whole-slide foundation model pre-trained with self-supervised and vision-language objectives. This process produces a feature matrix $H_j \in \mathbb{R}^{M\times d}$ for each WSI. To simplify notation, we drop the subscript j. To initialize semantic anchors, we apply K-means clustering on the training instances, obtaining cluster centers $S = \{s_k\}_{k=1}^{K}$. These centers serve as semantic anchors representing distinctive tissue patterns.

2.3 Context-Aware Clustering

Morphologically similar tissue types are often dispersed across distant anatomical regions, making it challenging to capture their contextual relationships and semantic consistency. MiCo addresses this challenge through its multi-layered context-aware clustering modules, which consist of two core components: the Cluster Route and the Cluster Reducer. The Cluster Route aggregates and propagates semantic information by linking instances of the same tissue type across distant regions, while the Cluster Reducer consolidates redundant anchors and facilitates information exchange between distinct semantic groups.

Cluster Route. Given a WSI containing M instances, each patch p_m is represented by a feature embedding $h_m \in \mathbb{R}^d$. To establish cross-regional intra-tissue correlations, we compute the cosine similarity between each instance feature h_i and $K^{'}$ learnable semantic anchors $\{s_k\}_{k=1}^{K^{'}}$:

$$A_{l(m,k)} = \frac{h_m^{\top} s_k}{|h_m| \cdot |s_k|}, \tag{1}$$

where $A_l \in \mathbb{R}^{M\times K^{'}}$ quantifies the semantic alignment between instances and semantic anchors.

The argmax operation ensures that each instance is assigned to its most relevant semantic anchor, preserving semantic consistency across pathological structures. However, argmax is non-differentiable, which disrupts gradient flow to the anchors. To address this, we use a straight-through estimator [19] to enable instance-to-anchor assignments while maintaining gradient propagation:

$$\hat{A}_l = \text{one-hot}\left(\arg\max(A_l)\right) + A_l - \text{sg}(A_l), \tag{2}$$

where $\text{sg}(\cdot)$ denotes the stop-gradient operator. During forward propagation, $\hat{A}_l \in \mathbb{R}^{M\times K^{'}}$ is a binary matrix where each row is one-hot encoded, assigning each instance to its most similar cluster. During backward propagation, gradients bypass the non-differentiable $\arg\max$ operation and flow through A_l, enabling end-to-end training.

For each semantic anchor s_k, we aggregate the features of its assigned instances to generate a context-aware representation:

$$\tilde{s}_k = \frac{1}{N_k} \sum_{m=1}^{M} \hat{A}_{l(m,k)} \cdot h_m, \quad N_k = \sum_{m=1}^{M} \hat{A}_{l(m,k)}. \tag{3}$$

where $\tilde{s}_k \in \mathbb{R}^d$ captures shared morphological semantics (e.g., tumor stroma or immune infiltration) within the anchor group, and N_k denotes the number of instances assigned to the k-th anchor. This aggregation step unifies spatially dispersed instances sharing similar semantics, directly mitigating spatial heterogeneity. We collect all $\tilde{s}_k$ into a matrix $\tilde{S} \in \mathbb{R}^{K^{'} \times d}$, which serves as the input to the Cluster Reducer module.

To propagate inter-region context, we align each instance feature h_m with its assigned semantic anchor semantics:

$$h_m^{'} = h_m + MLP(h_m + \hat{A}_{l(m,k)} \cdot \tilde{s}_k). \tag{4}$$

The updated feature $h_m^{'}$ combines instance-level features with cross-regional semantic context, creating a unified representation of spatially dispersed tissue structures. This enriched feature is then passed to the next context-aware clustering module for further refinement, enabling the model to progressively capture complex anatomical relationships.

Cluster Reducer. To reduce semantic redundancy, the Cluster Reducer merges anchors with similar characteristics by modeling inter-anchor relationships. Given initial semantic anchors $\tilde{S} \in \mathbb{R}^{K^{'} \times d}$, we transpose $\tilde{S}$ to $\tilde{S}^{\top} \in \mathbb{R}^{d \times K^{'}}$ and apply a MLP to model nonlinear interactions:

$$\tilde{S}^{'} = MLP(\tilde{S}^{\top}), \tag{5}$$

where $\tilde{S}^{'} \in \mathbb{R}^{d \times K^{'}/2}$ represents the refined anchors after redundancy reduction. The MLP captures semantic dependencies, such as merging scattered tumor-infiltrating lymphocyte clusters, by consolidating semantically similar anchors. Transposing $\tilde{S}^{'}$ back to $\tilde{S} \in \mathbb{R}^{K^{'}/2 \times d}$ yields compact and meaningful anchors that preserve essential tissue characteristics.

3 Experiments and Results

3.1 Datasets

Survival Prediction. We use seven publicly available cancer datasets from TCGA (**BLCA**, **BRCA**, **GBMLGG**, **HNSC**, **KIRC**, **KIRP**, and **LUAD**) in our experiments. These datasets collectively include data from 3,523 patients and 4,091 H&E diagnostic WSIs.

https://portal.gdc.cancer.gov

Table 1: Survival prediction performance comparison across seven cancer types.

Models	BLCA	BRCA	GBMLGG	HNSC	KIRC	KIRP	LUAD	MEAN
AMIL [5]	$0.551_{0.10}$	$0.576_{0.11}$	$0.736_{0.15}$	$0.590_{0.02}$	$0.681_{0.05}$	$0.778_{0.04}$	$0.584_{0.04}$	0.642
TransMIL [15]	$0.610_{0.07}$	$0.555_{0.04}$	$0.760_{0.14}$	$0.565_{0.02}$	$0.694_{0.05}$	$0.773_{0.07}$	$0.585_{0.09}$	0.649
DeepGraphConv [10]	$0.563_{0.08}$	$0.581_{0.10}$	$0.739_{0.14}$	$0.591_{0.02}$	$0.681_{0.02}$	$0.738_{0.04}$	$0.581_{0.02}$	0.639
PatchGCN [1]	$0.607_{0.05}$	$0.600_{0.12}$	$0.736_{0.16}$	$0.584_{0.01}$	$0.694_{0.05}$	$0.783_{0.05}$	$0.582_{0.04}$	0.655
ILRA [20]	$0.612_{0.04}$	$0.551_{0.07}$	$0.722_{0.16}$	$0.600_{0.01}$	$0.663_{0.04}$	$0.688_{0.05}$	$0.590_{0.04}$	0.632
HVTSurv [16]	$0.606_{0.05}$	$0.579_{0.09}$	$0.792_{0.01}$	$0.575_{0.03}$	$0.692_{0.04}$	$0.778_{0.06}$	$0.591_{0.03}$	0.658
RRTMIL [18]	$0.588_{0.05}$	$0.542_{0.08}$	$0.738_{0.16}$	$0.597_{0.02}$	$0.699_{0.04}$	$0.770_{0.06}$	$0.579_{0.04}$	0.644
WiKG [7]	$0.607_{0.07}$	$0.582_{0.09}$	$0.752_{0.16}$	$0.579_{0.02}$	$0.704_{0.06}$	$0.763_{0.09}$	$0.595_{0.06}$	0.654
MiCo	$\mathbf{0.619}_{0.03}$	$\mathbf{0.608}_{0.12}$	$\mathbf{0.813}_{0.02}$	$\mathbf{0.605}_{0.01}$	$\mathbf{0.715}_{0.05}$	$\mathbf{0.800}_{0.07}$	$\mathbf{0.604}_{0.04}$	**0.680**

Table 2: Cancer subtyping performance comparison across two cancer types.

Method	TCGA-BRCA			TCGA-NLCSC			MEAN		
	ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC
AMIL [5]	$0.909_{0.01}$	$0.858_{0.02}$	$0.948_{0.01}$	$0.923_{0.02}$	$0.923_{0.02}$	$0.979_{0.01}$	0.916	0.890	0.964
TransMIL [15]	$0.906_{0.01}$	$0.857_{0.02}$	$0.945_{0.02}$	$0.919_{0.02}$	$0.919_{0.02}$	$0.977_{0.02}$	0.913	0.889	0.961
DeepGraphConv [10]	$0.913_{0.01}$	$0.859_{0.01}$	$0.947_{0.01}$	$0.913_{0.02}$	$0.913_{0.02}$	$0.969_{0.02}$	0.913	0.886	0.958
PatchGCN [1]	$0.917_{0.01}$	$0.871_{0.02}$	$0.944_{0.02}$	$0.919_{0.02}$	$0.919_{0.02}$	$0.979_{0.01}$	0.918	0.895	0.962
ILRA [20]	$0.899_{0.01}$	$0.836_{0.03}$	$0.931_{0.02}$	$0.918_{0.02}$	$0.918_{0.02}$	$0.973_{0.02}$	0.909	0.885	0.952
RRTMIL [18]	$0.910_{0.01}$	$0.855_{0.01}$	$0.945_{0.02}$	$0.915_{0.02}$	$0.915_{0.02}$	$0.976_{0.01}$	0.913	0.890	0.961
WiKG [7]	$0.915_{0.02}$	$0.859_{0.03}$	$0.948_{0.01}$	$0.920_{0.03}$	$0.920_{0.03}$	$0.980_{0.01}$	0.917	0.889	0.964
MiCo	$\mathbf{0.922}_{0.01}$	$\mathbf{0.875}_{0.02}$	$\mathbf{0.952}_{0.01}$	$\mathbf{0.931}_{0.02}$	$\mathbf{0.931}_{0.02}$	$\mathbf{0.981}_{0.01}$	**0.927**	**0.903**	**0.967**

Cancer Subtyping. We conduct comparative experiments on two challenging public datasets: **TCGA-BRCA** and **TCGA-NSCLC**. The TCGA-BRCA dataset contains 1,034 H&E slides of two invasive cancer subtypes: invasive ductal carcinoma (IDC) and invasive lobular carcinoma (ILC). The TCGA-NSCLC dataset includes 1,030 H&E slides from two subtypes: Lung Squamous Cell Carcinoma (TCGA-LUSC) and Lung Adenocarcinoma (TCGA-LUAD).

3.2 Implementation Details

We compare MiCo with eight state-of-the-art methods: AMIL [5], TransMIL [15], DeepGraphConv [10], Patch-GCN [1], ILRA [20], HVTSurv [16], RRTMIL [18], and WiKG [7]. WSIs are preprocessed using CONCH v1.5 [4] to extract 448×448 patch features at 20× magnification. All experiments adopt 4-fold cross-validation, with datasets split into training, validation, and test sets at a ratio of 60:15:25. The semantic anchor number is set to 64. Training configurations are unified across methods: 200 epochs, batch size 1, learning rate 2e-4, and early stop 8. Survival prediction performance is evaluated using the Concordance Index (C-Index) with standard deviation, while cancer subtyping is assessed via Accuracy (ACC), F1-score (F1), and Area Under the Curve (AUC) metrics.

3.3 Results and Discussion

Survival Prediction. The experimental results demonstrate the superior performance of MiCo across seven cancer types for survival prediction tasks. As

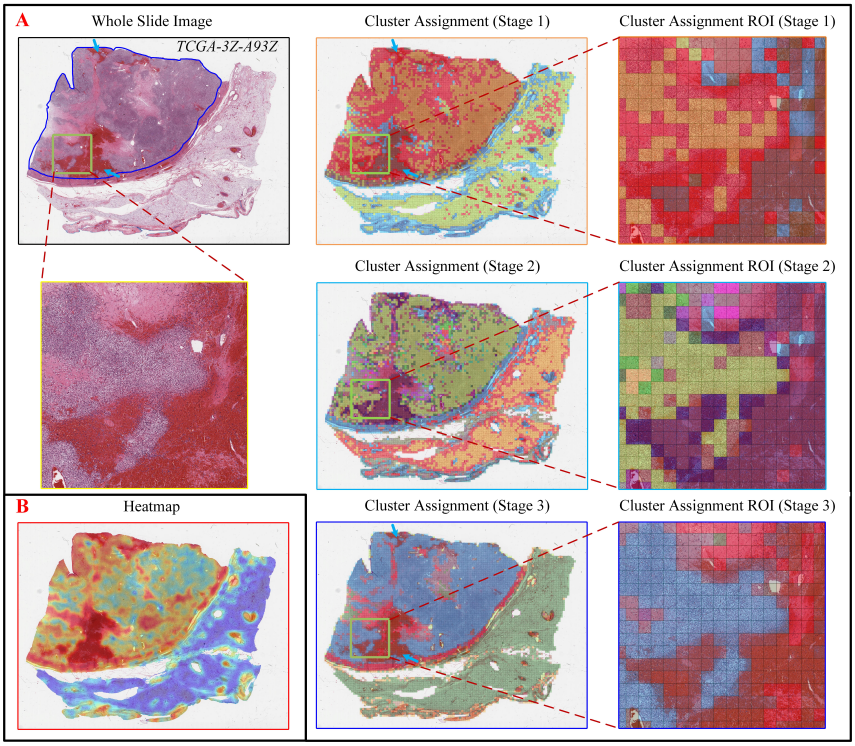

Fig. 2: **A.** Qualitative results of cluster assignment, showcasing WSI thumbnails, cluster assignment results, and corresponding regions of interest (ROIs) at different stages. Blue outlines denote ground truth tumor regions, while areas assigned to the same cluster are displayed in identical colors. The green arrows highlight how spatially distant regions are progressively aggregated into a single, semantically consistent cluster. **B.** Interpretability analysis of MiCo, with red regions in the heatmaps indicating areas of high attention.

shown in Table 1, MiCo achieves a mean C-index of 0.680, outperforming all other methods. This performance underscores the effectiveness of MiCo's context-aware clustering in addressing the challenges posed by spatial heterogeneity in histopathological WSIs. MiCo's ability to model cross-regional dependencies and enhance semantic associations across dispersed tissue regions enables it to capture complex morphological patterns and improve survival prediction accuracy.

Cancer Subtyping. The experimental results for cancer subtyping, as presented in Table 2, demonstrate MiCo's superior performance across two main datasets, TCGA-BRCA and TCGA-NLCSC. MiCo achieves state-of-the-art results in terms of accuracy, F1 score, and AUC, outperforming all other methods. With a mean ACC of 0.927, MiCo significantly surpasses the best-performing baseline, PatchGCN, which achieves a mean ACC of 0.918. These results highlight MiCo's ability to accurately classify cancer subtypes by leveraging its

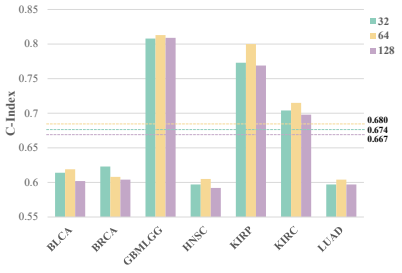

Fig. 3: Analysis of semantic anchors' impact.

Table 3: Average results of the ablation study for MiCo.

Method	Survival Prediction	Cancer Subtyping		
	C-Index	ACC	F1	AUC
w/o Semantic Anchors Init	0.668	0.922	0.896	0.964
w/o CluReducer	0.659	0.921	0.897	0.965
w/o CluRoute	0.656	0.910	0.890	0.955
MiCo	**0.680**	**0.927**	**0.903**	**0.967**

context-aware clustering mechanism and dynamic semantic anchors, effectively alleviating the challenges of spatial heterogeneity and semantic fragmentation in histopathological WSIs.

Ablation Study. To further determine the efficacy of each crucial component in MiCo, we conduct a series of ablation studies on survival prediction and cancer subtyping tasks. The average results, as summarized in Table 3, underscore the critical importance of each module in enhancing the overall performance of MiCo. The ablation experiments evaluate the impact of removing key components, including semantic anchor initialization, the CluReducer, and the CluRoute. Removing semantic anchor initialization reduces the C-index to 0.668, underscoring its importance in capturing discriminative morphological patterns. Omitting the CluReducer further decreases the C-index to 0.659, highlighting its role in consolidating redundant anchors and enhancing inter-anchor information exchange. Similarly, removing the CluRoute results in a C-index of 0.656, demonstrating its effectiveness in establishing cross-regional semantic relationships and refining instance-level representations.

Interpretability Analysis. As shown in Fig.2, we analyze MiCo's interpretability through cluster assignments and heatmaps. Fig.2.A demonstrates that cluster assignments of semantically related regions are progressively merged across stages, preventing overly coarse representations in later phases. The Cluster Reducer module halves cluster numbers at each stage of the MiCo.This progressive reduction aims to intelligently refine and consolidate semantic information,rather than causing coarseness or indiscriminately mixing distinct regional information. The heatmap results further reveal strong alignment with ground truth tumor annotations in Fig.2.B.

Semantic Anchors Analysis. As shown in Fig.3, we evaluate MiCo's survival prediction performance with different numbers of semantic anchors (32, 64, and 128). The mean C-index values are 0.674 (32 anchors), 0.680 (64 anchors), and 0.667 (128 anchors). The 64-anchor configuration achieves the highest mean C-index, balancing morphological diversity and redundancy.

4 Conclusion

In this paper, we propose MiCo, a novel Multiple Instance Learning framework with Context-Aware Clustering, designed to tackle spatial heterogeneity in histopathological WSIs. MiCo incorporates a Cluster Route module to strengthen cross-regional intra-tissue correlations and a Cluster Reducer module to consolidate redundant anchors while promoting information exchange between distinct semantic groups. Extensive experiments on nine cancer datasets validate MiCo's superiority over state-of-the-art methods in both survival prediction and cancer subtyping tasks.

Acknowledgement. This work was supported in part by Xinjiang Uygur Autonomous Region Key R&D program (No. 2024B03039-3), the National Natural Science Foundation of China (No. U24A20256), the Scientific Research Fund of Hunan Provincial Education Department (No. 23A0020), the Science and Technology Major Project of Changsha (No. kh2502004), the central government guides local funds for scientific and technological development of China (No. ZYYD2025QY25), the Central South University Innovation-Driven Research Programme (No. 2023CXQD018), the Fundamental Research Funds for the Central Universities of Central South University (No. 2024ZZTS0108), and High-Performance Computing Center of Central South University

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

References

1. Chen, R.J., Lu, M.Y., Shaban, M., Chen, C., Chen, T.Y., Williamson, D.F., Mahmood, F.: Whole slide images are 2d point clouds: Context-aware survival prediction using patch-based graph convolutional networks. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII 24. pp. 339–349. Springer (2021)
2. Cruz-Roa, A., Basavanhally, A., González, F., Gilmore, H., Feldman, M., Ganesan, S., Shih, N., Tomaszewski, J., Madabhushi, A.: Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. In: Medical imaging 2014: Digital pathology. vol. 9041, p. 904103. SPIE (2014)
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. Ding, T., Wagner, S.J., Song, A.H., Chen, R.J., Lu, M.Y., Zhang, A., Vaidya, A.J., Jaume, G., Shaban, M., Kim, A., Williamson, D.F.K., Chen, B., Almagro-Perez, C., Doucet, P., Sahai, S., Chen, C., Komura, D., Kawabe, A., Ishikawa, S., Gerber, G., Peng, T., Le, L.P., Mahmood, F.: Multimodal whole slide foundation model for pathology (2024), https://arxiv.org/abs/2411.19666
5. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International conference on machine learning. pp. 2127–2136. PMLR (2018)
6. Lee, M.: Recent advancements in deep learning using whole slide imaging for cancer prognosis. Bioengineering **10**(8), 897 (2023)

7. Li, J., Chen, Y., Chu, H., Sun, Q., Guan, T., Han, A., He, Y.: Dynamic graph representation with knowledge-aware attention for histopathology whole slide image analysis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11323–11332 (2024)
8. Li, J., Kuang, H., Liu, J., Yue, H., Wang, J.: Ca2cl: Cluster-aware adversarial contrastive learning for pathological image analysis. IEEE Journal of Biomedical and Health Informatics (2025)
9. Li, J., Liu, J., Yue, H., Cheng, J., Kuang, H., Bai, H., Wang, Y., Wang, J.: Darc: Deep adaptive regularized clustering for histopathological image classification. Medical image analysis **80**, 102521 (2022)
10. Li, R., Yao, J., Zhu, X., Li, Y., Huang, J.: Graph cnn for survival analysis on whole slide pathological images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 174–182. Springer (2018)
11. Lia, M., Horn, L.C., Sodeikat, P., Höckel, M., Aktas, B., Wolf, B.: The diagnostic value of core needle biopsy in cervical cancer: A retrospective analysis. Plos one **17**(1), e0262257 (2022)
12. Lipkova, J., Chen, T.Y., Lu, M.Y., Chen, R.J., Shady, M., Williams, M., Wang, J., Noor, Z., Mitchell, R.N., Turan, M., et al.: Deep learning-enabled assessment of cardiac allograft rejection from endomyocardial biopsies. Nature medicine **28**(3), 575–582 (2022)
13. Medvedev, O., Hedesiu, M., Ciurea, A., Lenghel, M., Rotar, H., Dinu, C., Roman, R., Termure, D., Csutak, C.: Perineural spread in head-and-neck malignancies: Imaging findings–an updated literature review. Bosnian Journal of Basic Medical Sciences **22**(1), 22 (2021)
14. Qu, L., Liu, S., Liu, X., Wang, M., Song, Z.: Towards label-efficient automatic diagnosis and analysis: a comprehensive survey of advanced deep learning-based weakly-supervised, semi-supervised and self-supervised techniques in histopathological image analysis. Physics in Medicine & Biology **67**(20), 20TR01 (2022)
15. Shao, Z., Bian, H., Chen, Y., Wang, Y., Zhang, J., Ji, X., et al.: Transmil: Transformer based correlated multiple instance learning for whole slide image classification. Advances in neural information processing systems **34**, 2136–2147 (2021)
16. Shao, Z., Chen, Y., Bian, H., Zhang, J., Liu, G., Zhang, Y.: Hvtsurv: hierarchical vision transformer for patient-level survival prediction from whole slide image. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 2209–2217 (2023)
17. Song, C., Hui, C., Lin, Q., Zhang, W., Li, S., Zhang, S., Zhu, H., Li, Z., Liu, S., Jiang, F., Li, X.: Lvpnet: A latent-variable-based prediction-driven end-to-end framework for lossless compression of medical images (2025), https://arxiv.org/abs/2506.17983
18. 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)
19. Van Den Oord, A., Vinyals, O., et al.: Neural discrete representation learning. Advances in neural information processing systems **30** (2017)
20. 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)
21. Zarella, M.D., Bowman, D., Aeffner, F., Farahani, N., Xthona, A., Absar, S.F., Parwani, A., Bui, M., Hartman, D.J.: A practical guide to whole slide imaging:

a white paper from the digital pathology association. Archives of pathology & laboratory medicine **143**(2), 222–234 (2019)