MicroMIL: Graph-Based Multiple Instance Learning for Context-Aware Diagnosis with Microscopic Images

JongWoo Kim[1*], Bryan Wong[1*], Huazhu Fu[2], Willmer Rafell Quiñones Robles[1], Young Sin Ko[3], and Mun Yong Yi[1†]

[1] Korea Advanced Institute of Science and Technology, Daejeon, South Korea
{gsds4885, bryan.wong, munyi}@kaist.ac.kr
[2] Institute of High Performance Computing, Agency for Science, Technology and Research (A*STAR), Singapore
[3] Seegene Medical Foundation, Pathology Center, Seoul, South Korea

Abstract. Cancer diagnosis has greatly benefited from the integration of whole-slide images (WSIs) with multiple instance learning (MIL), enabling high-resolution analysis of tissue morphology. Graph-based MIL (GNN-MIL) approaches have emerged as powerful solutions for capturing contextual information in WSIs, thereby improving diagnostic accuracy. However, WSIs require significant computational and infrastructural resources, limiting accessibility in resource-constrained settings. Conventional light microscopes offer a cost-effective alternative, but applying GNN-MIL to such data is challenging due to extensive redundant images and missing spatial coordinates, which hinder contextual learning. To address these issues, we introduce **MicroMIL**, the first weakly-supervised MIL framework specifically designed for images acquired from conventional light microscopes. MicroMIL leverages a representative image extractor (RIE) that employs deep cluster embedding (DCE) and hard Gumbel-Softmax to dynamically reduce redundancy and select representative images. These images serve as graph nodes, with edges computed via cosine similarity, eliminating the need for spatial coordinates while preserving contextual information. Extensive experiments on a real-world colon cancer dataset and the BreakHis dataset demonstrate that MicroMIL achieves state-of-the-art performance, improving both diagnostic accuracy and robustness to redundancy. The code is available at https://github.com/kimjongwoo-cell/MicroMIL

Keywords: Digital Pathology · Conventional Light Microscopes · Microscopy Images · Multiple Instance Learning

1 Introduction

Cancer remains a leading global cause of mortality, necessitating advancements in diagnostic technologies to enhance early detection and improve survival rates.

[*] Co-first authors with equal contribution.

[†] Corresponding author.

Whole-slide imaging (WSI) has emerged as a transformative tool in digital pathology, offering high-resolution insights into tissue morphology and disease-related anomalies [6, 13]. However, WSIs require substantial infrastructure due to their high acquisition costs, memory demands, and lengthy processing times, making them less practical in resource-limited settings [23, 2].

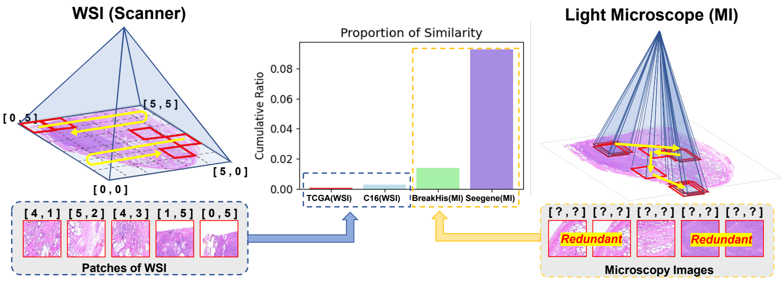

Fig. 1. Left: Valid patches from WSI (scanner) are acquired using a sliding-window approach and have absolute positions. **Right:** Light microscopy images lack known positions and contain many redundancies due to subjective capture by pathologists. **Middle:** Ratio of image pairs exceeding the redundancy threshold (0.995).

While WSI scanners are becoming more accessible, light microscopes remain far more widely used, especially in low-resource settings. Microscopy images thus offer a practical and cost-effective alternative for enabling AI-driven diagnostic solutions in diverse healthcare contexts [21]. Low-cost optical microscopes tailored for low- and middle-income countries continue to be developed and microscopy-based diagnostics remain the clinical standard [19, 26].

Despite these advantages, as illustrated in Figure 1, light microscopy images pose unique challenges, including the **absence of absolute spatial coordinates** due to manual acquisition by pathologists and **significant redundancy** caused by multiple image captures. To highlight these issues, we compare microscopic and WSI datasets (TCGA NSCLC[4] and Camelyon16 [3]) and show that light microscopy images exhibit significantly higher redundancy, with the highest observed in real-world microscopic datasets.

Recent advancements in weakly-supervised multiple instance learning (MIL) have facilitated the use of WSIs for cancer diagnosis by requiring only slide-level labels, thereby reducing the need for exhaustive annotations [7]. Within this paradigm, graph-based MIL (GNN-MIL) models [5, 4] have shown promise by leveraging spatial relationships among patches to capture contextual information, representing patches as nodes and their interactions as edges. However, these methods are inherently designed for WSIs and cannot be directly applied to light microscopy images, where spatial coordinates are unknown and image

[4] https://www.cancer.gov/tcga

redundancy is prevalent. Overcoming these limitations requires a specific approach that accommodates the unique characteristics of light microscopy images while preserving the benefits of graph-based contextual modeling.

To address the challenges of absent spatial coordinates and high redundancy, we propose **MicroMIL**, the first weakly-supervised MIL framework specifically designed for conventional light microscopy images. MicroMIL introduces a representative image extractor (RIE) that leverages deep cluster embedding (DCE) [11] to dynamically group redundant images and hard Gumbel-Softmax [14] to select a representative image per cluster. These selected images serve as graph nodes, with edges formed using cosine similarity to capture contextual information among instances. While prior approaches have relied on statistical heuristics or ensemble-based methods [20, 8], the most related method [15] does not address critical challenges specific to light microscopy images. MicroMIL is explicitly designed to overcome these limitations through a graph-based formulation that operates without relying on spatial metadata.

By enabling end-to-end representative feature selection, MicroMIL jointly optimizes clustering and instance selection within a unified framework, ensuring that the most informative representations contribute to the final prediction. To achieve this, we propose an online redundancy-aware learning strategy that dynamically refines instance selection while maintaining feature diversity. The graph-based representation further enhances structural preservation by connecting similar nodes, mitigating the loss of spatial information. Extensive experiments on a real-world colon cancer dataset and the BreakHis dataset validate MicroMIL's effectiveness, demonstrating significant gains in diagnostic accuracy and redundancy robustness compared to state-of-the-art MIL methods.

2 Methodology

Each patient is treated as a *bag*, and the corresponding light-microscope images as *instances*, following the MIL paradigm. The goal is to predict patient diagnosis in a weakly-supervised setting without instance-level labels. We propose MicroMIL, a MIL framework for microscopic image analysis (see Figure 2). The framework consists of three components: (1) a frozen pre-trained feature extractor for generating image features, (2) a RIE that reduces redundancy by clustering similar images using DCE and selecting representatives via hard Gumbel-Softmax, and (3) a graph-based aggregation module, where nodes represent the selected represented images and edges are constructed using an upper triangular cosine similarity matrix to link similar nodes. Finally, a GNN captures contextual information for analysis.

2.1 Representative Image Extractor

In the embedding-based MIL framework, a frozen pre-trained feature extractor E maps each microscopic image I_s to a d-dimensional feature vector $f_s = E(I_s)$, forming the feature set $F = \{f_1, f_2, \dots, f_S\}$, where S varies across patients. To

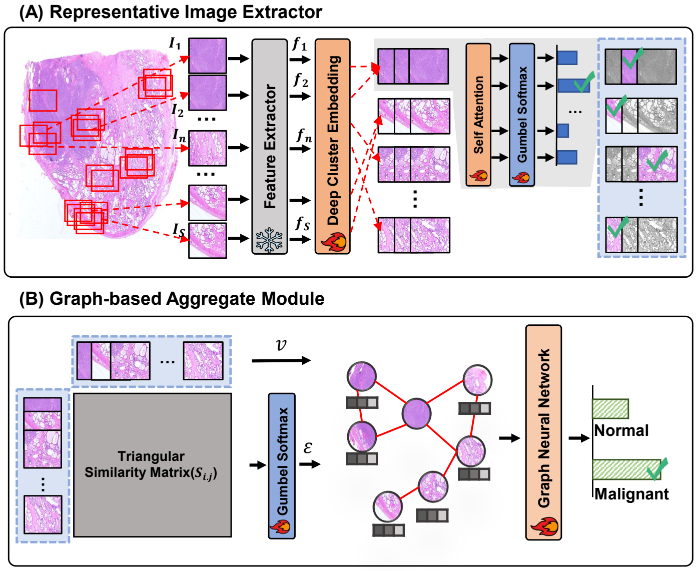

Fig. 2. Proposed end-to-end MicroMIL framework.

group redundant images, we employ deep cluster embedding (DCE) [25], which iteratively assigns data points to clusters and refines cluster centers to minimize intra-cluster distances. Let $\mu_c \in \mathbb{R}^d$ be the centroid of the c-th cluster, where $c = 1, \ldots, C$. The soft assignment probability $z_{s,c}$, indicating the likelihood of f_s belonging to the c-th cluster, is defined as:

$$z_{s,c} = \frac{\left(1 + \|f_s - \mu_c\|^2\right)^{-1}}{\sum_{j=1}^{C} \left(1 + \|f_s - \mu_j\|^2\right)^{-1}}, \quad Z \in \mathbb{R}^{S \times C} \tag{1}$$

The DCE algorithm alternates between updating the centroids μ_c and refining cluster assignments Z until convergence.

To select the most representative feature from each cluster, we use the hard Gumbel-Softmax function [14], which allows for end-to-end differentiability. Given logits X_x and Gumbel noise $g_x \sim \text{Gumbel}(0, 1)$, the hard Gumbel-Softmax function is defined as:

$$\text{HardGumbel}(X) = \text{one\_hot}\left(\arg\max_x (X_x + g_x)\right) \tag{2}$$

Applying this function to feature-cluster interactions, the hard cluster assignments $\tilde{Z}$ are determined as:

$$\tilde{z}_{s,c} = \text{HardGumbel}(s_{s,c}), \quad \tilde{Z} \in \mathbb{R}^{S \times C} \tag{3}$$

where $s_{s,c} = w^\top (f_s \odot z_{:,c})$, with $w \in \mathbb{R}^d$ being a learnable weight vector and $\odot$ denoting element-wise multiplication.

The representative feature q_c of cluster c is computed as follows:

$$q_c = \sum_{s=1}^{S} \tilde{z}_{s,c} f_s, \quad Q \in \mathbb{R}^{C \times d} \tag{4}$$

This process combines DCE for clustering and the hard Gumbel-Softmax for selecting representative features, ensuring inter-cluster separation and intra-cluster compactness. By focusing on representative features, this approach improves subsequent classification performance while reducing redundancy.

2.2 Graph-based Aggregate Module

To model relationships among clusters, we construct a graph G, where nodes represent representative feature clusters and edges capture pairwise similarities. Given representative cluster embeddings $Q = \{q_1, q_2, \ldots, q_C\}$, where $q_c \in \mathbb{R}^d$, the pairwise similarity is computed using cosine similarity as $S_{ij} = \frac{q_i^\top q_j}{\|q_i\|\|q_j\|}$, with $\|q_i\|$ denoting the Euclidean norm of q_i. A value of S_{ij} closer to 1 indicates higher similarity between clusters.

To retain only the most important relationships, we apply the same hard Gumbel-Softmax function to the similarity matrix:

$$\tilde{m}_{i,j} = \text{HardGumbel}(S_{ij}), \quad \tilde{M} \in \mathbb{R}^{C \times C} \tag{5}$$

The resulting graph $G = (V, E)$ is defined by nodes $V = \{1, 2, \ldots, C\}$ and edges $E = \{(i, j) \mid \tilde{m}_{i,j} > 0\}$.

Once the graph is constructed, a GNN propagates and refines the cluster embeddings. The initial node features are $H^{(0)} = R$ and through L GNN layers, node embeddings are updated by aggregating information from neighboring nodes. The entire process is represented as:

$$y = \sigma\left(W_{\text{class}} \cdot \text{mean}\left(\text{GNN}(G, R)\right)\right) \tag{6}$$

where $\text{GNN}(G, R)$ represents the L-layer operations on G, W_{class} is the classification weight matrix, $\text{mean}(\cdot)$ aggregates node embeddings, and σ is the activation function. The entire framework, including DCE, RIE, and GNN, is trained end-to-end using binary cross-entropy (BCE) loss.

3 Experiments and Results

3.1 Datasets

BreakHis BreakHis [24] is a widely used benchmark dataset for microscopy image analysis and cancer research. It comprises 7,909 images from 81 patients, with an average of 96.4 images per patient. Among these, 2,480 images (from 24 patients) are labeled as normal, while 5,429 images (from 57 patients) are labeled as malignant. The dataset spans multiple magnifications (40×, 100×, 200×, and 400×) and is collected under controlled conditions, which may limit its applicability to real-world scenarios.

Real-world (Seegene) Dataset Our real-world dataset, collected from the Seegene Medical Foundation, consists of 135,100 images from 899 patients, averaging 150.3 images per patient. The dataset was approved by both the foundation's IRB (SMF-IRB-2020-007) and the KAIST IRB (KAIST-IRB-23-214). It includes 52,339 images (from 493 patients) labeled as normal and 82,761 images (from 406 patients) labeled as malignant, primarily at 100× and 200× magnifications, aligning with common clinical practices.

3.2 Implementation Details

For a fair comparison, we use ResNet18 pre-trained on ImageNet as the feature extractor for all models, with a hidden dimension of 128 for consistency. MicroMIL is tuned with a dropout rate of 0.5, a learning rate of 1×10^{-3}, and the Adam optimizer. For the final results, we use 36 clusters for the real-world (Seegene) dataset and 16 for BreakHis, given its smaller image count per patient. We experiment with 16 (4^2), 25 (5^2), and 36 (6^2) clusters and report the best-performing configuration. Performance differences across these settings are minimal and MicroMIL consistently outperforms all baselines. While online clustering (DCE) requires a predefined cluster count, we aim to explore automatic cluster number selection in future work. All models, implemented in PyTorch, are trained with two graph layers on an NVIDIA GeForce RTX 3080 GPU.

3.3 Baselines Comparison

Table 1. Performance metrics of baselines and MicroMIL on Real-world (Seegene) and BreakHis datasets. Best results are in **bold** and second-best results are underlined.

	Real-world (Seegene)			**BreakHis**		
Model	**ACC**	**AUC**	**F1**	**ACC**	**AUC**	**F1**
ABMIL [ICML'18] [12]	0.9444	0.9764	0.9433	0.8929	0.8947	0.8805
MS-DA-MIL [CVPR'20] [10]	0.9556	0.9829	0.9514	0.8929	0.9591	0.9268
DSMIL [CVPR'21] [16]	0.9444	0.9829	0.9440	0.8214	0.8947	0.8155
CLAM [Nat BioMed'21] [18]	0.9556	0.9873	0.9552	0.9286	0.9298	0.9181
TransMIL [NeurIPS'21] [22]	0.9778	0.9873	0.9776	0.8929	0.9825	0.9268
DTFD-MIL [CVPR'22] [27]	0.9611	0.9901	0.9607	0.9286	0.9766	0.9222
IBMIL [CVPR'23] [17]	0.9611	0.9894	0.9606	0.9286	0.9532	0.9181
ACMIL [ECCV'24] [28]	0.9611	0.9893	0.9606	0.8929	0.9474	0.8857
MicroMIL	**0.9922**	**0.9994**	**0.9925**	**0.9643**	**0.9942**	**0.9730**

Table 1 shows that MicroMIL consistently surpasses baseline MIL models across all evaluated metrics on both real-world and public datasets. This performance advantage stems from the MicroMIL's specific design to address two key challenges of light microscopy datasets: image redundancy and no absolute position. In contrast, existing MIL models are designed for WSIs obtained from scanners, thereby without the need to account for these characteristics. By effectively tackling the unique characteristics present, MicroMIL proves to be exceptionally well-suited for patient diagnosis using light microscopy images.

3.4 Impact of Representative Image Extractor

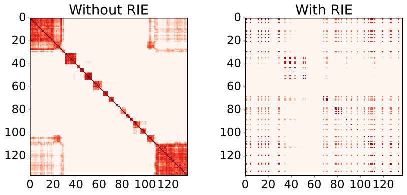

Fig. 3. Similarity-based edge formation probabilities ($\geq$ 0.8) heatmaps for real-world (Seegene) data, comparing cases *without (Left)* and *with (Right)* the RIE.

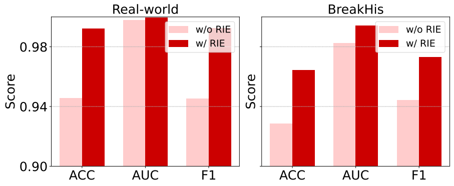

Fig. 4. Performance metrics *without (w/o)* and *with (w/)* the RIE on Real-world (Seegene) and BreakHis datasets.

In histopathology, patient-level prediction requires information exchange across diverse tissue regions [1, 9], but redundant images hinder this process. To address this, we propose the Representative Image Extractor (RIE), which selects representative images from visually similar clusters to enhance diversity and improve predictive performance. Figure 3 (*without* RIE) shows that redundancy limits diverse interactions and leads to performance degradation (Figure 4), which RIE successfully mitigates by filtering redundant instances. The effect varies depending on dataset redundancy, with smaller improvement in BreakHis due to lower redundancy, whereas the real-world dataset shows a larger improvement, demonstrating RIE's effectiveness in handling highly redundant data.

3.5 Connecting Relative Neighborhood Nodes Method

Light microscopy images lack spatial information, necessitating effective edge construction. We utilize a GNN-based method that connects nodes to their most similar neighbors. As shown in Figure 5, performance drops without connections due to a lack of contextual relationships. Random connections lead to weak interactions, while cosine similarity-based edges, linking highly similar nodes, capture meaningful relationships and outperform random or reverse-similarity (1/similarity) edges. This highlights the importance of leveraging similarity to enhance context integration and predictive performance.

3.6 Ablation Studies

Representative Image Extractor Methods. To enhance feature learning, we adopt an online approach using DCE and Gumbel-Softmax for selecting influential representative images. As shown in Figure 6, a comparison of clustering

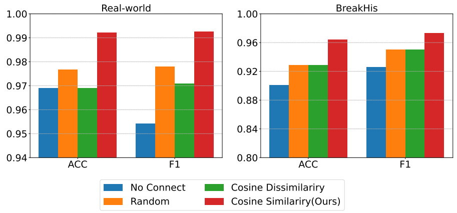

Fig. 5. Performance metrics of different edge generation methods

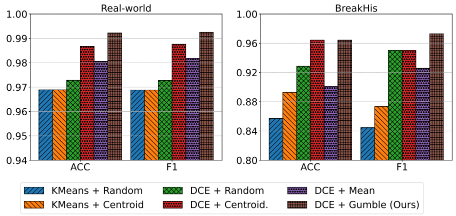

Fig. 6. Performance metrics of different representative image extractor methods

methods (KMeans, DCE) and selection strategies (Random, Mean, Centroid, Gumbel-Softmax) reveals that offline KMeans underperforms, while online DCE + Gumbel-Softmax achieves superior results. Notably, Gumbel-Softmax outperforms Mean and Centroid by better highlighting key images, demonstrating the effectiveness of online clustering and selection in optimizing feature learning. This online approach enables dynamic adaptation, ensuring robust feature representations and reducing the risk of suboptimal cluster assignments.

Table 2. Robustness to redundancy in baseline models and MicroMIL. Arrows indicate data flow direction (train → test). We count images exceeding the 0.995 redundancy threshold (Figure 1, middle) per patient, then select the top 10% (**T10**) and bottom 10% (**B10**) of patients. Best results are in **bold**.

	(1) T10 → B10			**(2) B10 → T10**			**(3) T10 → T10**		
Model	**ACC**	**AUC**	**F1**	**ACC**	**AUC**	**F1**	**ACC**	**AUC**	**F1**
ABMIL	0.8090	0.8592	0.7966	0.9213	0.9592	0.9210	0.9091	0.9229	0.9089
MS-DA-MIL	0.9101	0.9526	0.9248	0.9213	0.9642	0.9248	0.9091	0.9348	0.9209
DSMIL	0.9101	0.9474	0.9092	0.9326	0.9755	0.9319	0.9091	0.9438	0.9089
CLAM	0.9326	0.9796	0.9315	0.9213	0.9770	0.9194	0.9318	0.9521	0.9318
TransMIL	0.9213	0.9776	0.9212	0.8989	0.8526	0.8963	0.9318	0.8854	0.9309
DTFD-MIL	0.9438	0.9658	0.9428	0.9213	0.9750	0.9203	0.9318	0.9583	0.9318
IBMIL	0.9326	0.9709	0.9319	0.9326	0.9719	0.9319	0.9318	0.9458	0.9315
ACMIL	0.9434	0.9704	0.9431	0.9213	0.9689	0.9210	0.9318	0.9521	0.9315
MicroMIL	**0.9663**	**0.9842**	**0.9630**	**0.9551**	**0.9801**	**0.9542**	**0.9545**	**0.9958**	**0.9524**

Robustness on Image Redundancy Shift. To evaluate MicroMIL's robustness to image redundancy, we set a similarity threshold of 0.995 to identify redundant image pairs. Table 2 shows that baseline MIL methods degrade under extreme redundancy settings (B10→T10, T10→T10), while MicroMIL consistently maintains high performance. Even in low-redundancy simulations (T10→B10), MicroMIL outperforms all baselines. These results confirm MicroMIL's ability to extract critical features and remain effective in any scenario.

4 Conclusion

We introduce MicroMIL, the first weakly-supervised MIL framework specifically designed for conventional light microscopy images, addressing the limitations of GNN-MIL approaches that rely on spatial coordinates and exhibit low redundancy tolerance. By integrating deep cluster embedding (DCE) and hard Gumbel-Softmax, MicroMIL effectively reduces redundancy and selects representative instances, enabling a graph-based representation without requiring absolute spatial positioning while explicitly modeling contextual cues. Experiments on the real-world and BreakHis datasets demonstrate state-of-the-art performance, improving diagnostic accuracy while maintaining robustness to redundancy. MicroMIL offers a scalable, spatially agnostic solution that advances weakly-supervised MIL for microscopy imaging in resource-constrained settings.

Acknowledgments. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No.RS-2022-NR068758).

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

References

1. Ahn, S., Huang, H.: Multiregion sequence analysis to predict intratumor heterogeneity and clonal evolution. Deep Sequencing Data Analysis pp. 283–296 (2021)
2. Alkassar, S., Jebur, B.A., Abdullah, M.A., Al-Khalidy, J.H., Chambers, J.A.: Going deeper: a magnification-invariant approach for breast cancer classification using histopathological images. IET Computer Vision **15**(2), 151–164 (2021)
3. Bejnordi, B.E., Veta, M., Van Diest, P.J., Van Ginneken, B., Karssemeijer, N., Litjens, G., Van Der Laak, J.A., Hermsen, M., Manson, Q.F., Balkenhol, M., et al.: Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. Jama **318**(22), 2199–2210 (2017)
4. Bontempo, G., Bartolini, N., Lovino, M., Bolelli, F., Virtanen, A., Ficarra, E.: Enhancing pfi prediction with gds-mil: A graph-based dual stream mil approach. In: International Conference on Image Analysis and Processing. pp. 550–562. Springer (2023)
5. Chan, T.H., Cendra, F.J., Ma, L., Yin, G., Yu, L.: Histopathology whole slide image analysis with heterogeneous graph representation learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 15661–15670 (2023)
6. Evans, A.J., Krupinski, E.A., Weinstein, R.S., Pantanowitz, L.: 2014 american telemedicine association clinical guidelines for telepathology: Another important step in support of increased adoption of telepathology for patient care. Journal of Pathology Informatics **6** (2015)
7. Gadermayr, M., Tschuchnig, M.: Multiple instance learning for digital pathology: A review of the state-of-the-art, limitations & future potential. Computerized Medical Imaging and Graphics p. 102337 (2024)
8. Gandomkar, Z., Brennan, P.C., Mello-Thoms, C.: Mudern: Multi-category classification of breast hikumartopathological image using deep residual networks. Artificial intelligence in medicine **88**, 14–24 (2018)

9. Halperin, R.F., Liang, W.S., Kulkarni, S., Tassone, E.E., Adkins, J., Enriquez, D., Tran, N.L., Hank, N.C., Newell, J., Kodira, C., et al.: Leveraging spatial variation in tumor purity for improved somatic variant calling of archival tumor only samples. Frontiers in oncology **9**, 119 (2019)
10. Hashimoto, N., Fukushima, D., Koga, R., Takagi, Y., Ko, K., Kohno, K., Nakaguro, M., Nakamura, S., Hontani, H., Takeuchi, I.: Multi-scale domain-adversarial multiple-instance cnn for cancer subtype classification with unannotated histopathological images. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 3852–3861 (2020)
11. Huang, P., Huang, Y., Wang, W., Wang, L.: Deep embedding network for clustering. In: 2014 22nd International conference on pattern recognition. pp. 1532–1537. IEEE (2014)
12. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International conference on machine learning. pp. 2127–2136. PMLR (2018)
13. Iyengar, J.N.: Whole slide imaging: The futurescape of histopathology. Indian Journal of Pathology and Microbiology **64**(1), 8 (2021)
14. Jang, E., Gu, S., Poole, B.: Categorical reparameterization with gumbel-softmax. arXiv preprint arXiv:1611.01144 (2016)
15. Kim, J.W., Jeong, M.K., Park, H.M., Ko, Y.S., Yi, M.: Leveraging spatial relationships in microscopic images for patient cancer diagnosis. In: 2024 IEEE International Symposium on Biomedical Imaging (ISBI). pp. 1–5. IEEE (2024)
16. 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)
17. 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)
18. Lu, M.Y., Williamson, D.F., Chen, T.Y., Chen, R.J., Barbieri, M., Mahmood, F.: Data-efficient and weakly supervised computational pathology on whole-slide images. Nature biomedical engineering **5**(6), 555–570 (2021)
19. McDermott, S., Ayazi, F., Collins, J., Knapper, J., Stirling, J., Bowman, R., Cicuta, P.: Multi-modal microscopy imaging with the openflexure delta stage. Optics Express **30**(15), 26377–26395 (2022)
20. Nguyen, H.G., Blank, A., Dawson, H.E., Lugli, A., Zlobec, I.: Classification of colorectal tissue images from high throughput tissue microarrays by ensemble deep learning methods. Scientific reports **11**(1), 1–11 (2021)
21. Sangameswaran, R.: Maiscope: A low-cost portable microscope with built-in vision ai to automate microscopic diagnosis of diseases in remote rural settings. arXiv preprint arXiv:2208.06114 (2022)
22. 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)
23. Spanhol, F.A., Oliveira, L.S., Petitjean, C., Heutte, L.: Breast cancer histopathological image classification using convolutional neural networks. In: 2016 international joint conference on neural networks (IJCNN). pp. 2560–2567 (2016)
24. Spanhol, F.A., Oliveira, L.S., Petitjean, C., Heutte, L.: A dataset for breast cancer histopathological image classification. Ieee transactions on biomedical engineering **63**(7), 1455–1462 (2015)

25. Xie, J., Girshick, R., Farhadi, A.: Unsupervised deep embedding for clustering analysis. In: International conference on machine learning. pp. 478–487. PMLR (2016)
26. Zhang, H., Zhang, W., Zuo, Z., Yang, J.: Towards ultra-low-cost smartphone microscopy. Microscopy Research and Technique **87**(7), 1521–1533 (2024)
27. Zhang, H., Meng, Y., Zhao, Y., Qiao, Y., Yang, X., Coupland, S.E., Zheng, Y.: Dtfd-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)
28. 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)