Marginal Constrained Morphological Prototype Learning for Patch Search in Whole Slide Images

Sihyeon Park[1], Jungwoo Park[1], Hyunjae Kim[2], Jaewoo Kang[1,3,✉], and Bumsoo Kim[4,✉]

[1] Department of Computer Science, Korea University, Seoul, Republic of Korea
[2] Yale School of Medicine, Yale University, New Haven, CT, USA
[3] AIGEN Sciences Inc, Seoul, Republic of Korea kangj@korea.ac.kr
[4] School of CSE, Chung Ang University, Seoul, Republic of Korea bumsoo@cau.ac.kr

Abstract. Digital histopathology images are fundamental to modern oncology, enabling precise cancer diagnosis and the identification of clinically relevant morphological features. Due to the gigapixel scale of Whole Slide Images (WSIs), the dominant paradigm remains a two-step pipeline that encodes patches with a pre-trained backbone and aggregates features via Multi Instance Learning (MIL). However, this paradigm inherently limits optimization; encoders are often kept fixed, and prototype-based approaches are not always tightly coupled with the slide-level discriminative objective. Here, we propose Marginal-Constrained Morphological Prototype Learning (MMPL), a framework that formulates WSI diagnosis as the selective retrieval of informative patches. MMPL introduces two core innovations: (1) global prototype learning optimized via Sinkhorn-Knopp assignment against a momentum-stabilized feature queue (2) prototype-guided top-k patch selection. By focusing only on the top-k most informative patches, MMPL reduces the number of processed instances up to 99.3%, substantially lowering computational requirements. Experimental results on three WSI benchmarks demonstrate that MMPL outperforms the previous best baseline by up to 4.2% in weighted F1 for metastasis detection and 1.8% for tumor subtyping. Crucially, this breakthrough efficiency enables the vision encoder to learn discriminative morphological features, further rendering end-to-end training practical. The code will be available at https://github.com/dmis-lab/MMPL.

Keywords: Whole Slide Image · Multi Instance Learning · Global Prototype Learning.

1 Introduction

Whole-slide images (WSIs) are central to modern computational pathology, enabling cancer diagnosis [3, 5], prognosis, therapeutic response prediction [4, 19],

✉ : corresponding author

and discovery of clinically meaningful morphology [2, 25]. Yet, WSIs pose a fundamental learning challenge: they are gigapixel-scale images in which diagnostically decisive evidence often occupies only a small fraction of tissue. A robust slide-level model must therefore solve two problems simultaneously: it must identify the few morphologically informative regions among thousands of candidates, and it must do so under severe computational constraints that make exhaustive end-to-end processing impractical.

Current WSI pipelines [13, 14, 18] typically extract and encode patches using a pretrained vision backbone, and aggregate patch embeddings with multiple instance learning (MIL). While effective, this paradigm exposes a persistent trade-off between efficiency and diagnostic fidelity. Sampling-based methods [14, 29] reduce computation via subset selection, but their heuristic nature risks discarding rare, clinically critical morphology. Attention-based methods [13, 15, 21, 27, 28] learn soft importance weights over patches, but require processing all of them, thereby limiting scalability and end-to-end optimization. Prototype-based approaches [20, 22, 26] offer a promising alternative by summarizing redundant morphology into compact representative patterns. However, when prototypes are learned independently of the slide-level objective, they tend to capture frequent but non-discriminative structures, produce redundant concepts, or underrepresent critical regions. This observation motivates viewing prototype learning as a task-aware retrieval problem, where prototypes serve as semantic anchors for selecting informative patches under the supervision of the slide-level objective.

To this end, we propose MMPL, a unified framework that frames WSI slide-level prediction as a selective retrieval process, explicitly aligning patch selection and prototype learning with the global slide-level classification objective. MMPL learns a set of global morphological prototypes and assigns patches to prototypes through a task-aware optimal transport objective, encouraging balanced utilization while preserving morphological diversity. The resulting top-k retrieval mechanism selects a compact set of representative patches, reducing computation while retaining clinically salient evidence. Crucially, by coupling patch selection, prototype learning, and slide-level prediction within one objective, MMPL enables end-to-end optimization of the vision encoder.

Our experiments on three WSI benchmarks demonstrate the effectiveness of MMPL. With a ResNet50 backbone [12], our method achieves substantial improvements on two benchmarks, with gains of up to 4.2% in weighted F1 and 1.8% in AUC, while remaining competitive on the third. Furthermore, when integrated with the UNI backbone [8] model, MMPL outperforms the prototype-based baseline on most metrics, highlighting the benefit of aligning patch selection and prototype learning. Moreover, under end-to-end training, MMPL outperforms existing end-to-end MIL baseline.

2 Method

We present MMPL, a prototype-driven framework for WSI analysis that integrates optimal transport with dynamic patch retrieval (Figure 1).

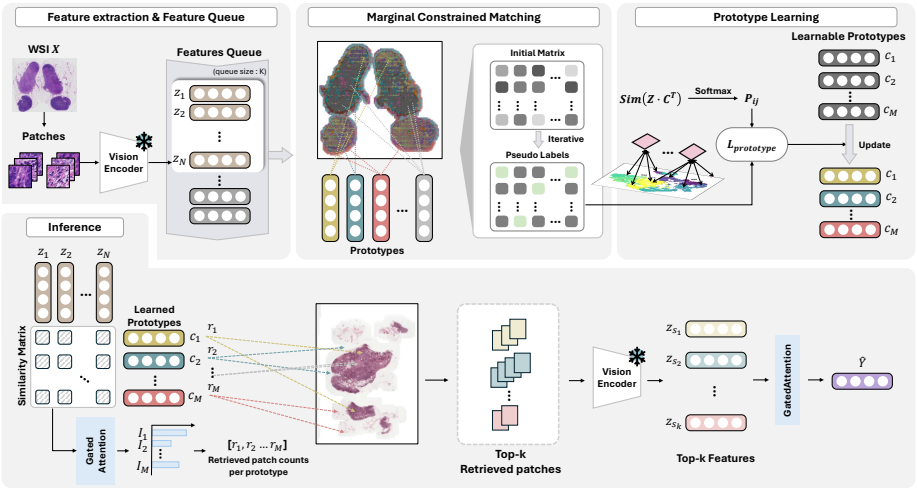

Fig. 1: Overview of the MMPL framework.

2.1 Sampling WSI patch instances

Given a WSI X, the slide-level prediction $\hat{Y}$ is generated by training a vision encoder $h(\cdot;\phi)$ and a feature-level classifier $g(\cdot;\theta)$. Due to the high resolution of WSIs, X is divided into N non-overlapping smaller instances $\{x_1,\ldots,x_N\} \subset \mathbb{R}^{W\times H\times 3}$ to encode them into $\mathcal{Z}=\{\mathbf{z}_1,\ldots,\mathbf{z}_N\}\subset\mathbb{R}^d$ where $z_i = h(x_i;\phi)$.

With WSI-level supervision, labels for each x_i are unavailable, so the problem is treated as multi-instance learning (MIL), with slide-level prediction given by:

$$\hat{Y} = g(\mathcal{Z};\theta) = \sigma\left(\sum_{i=1}^{N} a_i \mathbf{z}_i\right), \tag{1}$$

where g is an attention-based classifier with attention weights $\{a_1,\cdots,a_N\}$ [13].

Since encoding all patches is computationally intensive, using a subset of patches $\mathcal{S}\subset X$ is a practical solution, but naive selections such as random sampling have risks of essential region missing. Instead, we use global prototypes to guide the model toward diagnostically informative features.

2.2 Marginal Constrained Morphological Prototype Learning

To learn a set of global morphological prototypes, we propose a framework that combines an online prototype learning with a feature queue [11]. This dynamically aligns patch features to learnable prototypes in a way that facilitates stable and consistent update throughout training.

Feature Queue. To encourage global context sharing beyond the current WSI, we maintain a feature queue $\hat{\mathbf{Z}}=[\hat{\mathbf{z}}_1,\ldots,\hat{\mathbf{z}}_K]$ where $N\le K$, updated in a first-in-first-out (FIFO) manner. The queue stores features from both previous and

current slides, allowing the prototypes to reflect global morphological patterns rather than being biased toward the localized distribution of the current slide. In the end-to-end training setup, we instead use features from a momentum vision encoder [7, 9, 11] to prevent from representative shift via rapid change of $h(\cdot, \phi)$.

Prototype Learning via Sinkhorn-Knopp Algorithm. We formulate global prototype learning as the optimal transport (OT) problem. Given $\hat{\mathbf{Z}}$ and the prototypes $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_M]$, we find a nonnegative transport matrix $Q \in \mathbb{R}_+^{K \times M}$, where Q_{ij} denotes the assignment mass from $\hat{\mathbf{z}}_i$ to prototype $\mathbf{c}_j$. Let $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_N] \in \mathbb{R}^{N \times d}$ be in the current batch, then assignment predictions are:

$$P_{ij} = \frac{\exp\left(\mathbf{z}_i^T \mathbf{c}_j / \tau\right)}{\sum_{k=1}^{M} \exp\left(\mathbf{z}_i^T \mathbf{c}_k / \tau\right)}, \tag{2}$$

where τ is a temperature that we set to 0.1 for all experiments. All vectors, including $\mathbf{z}_i$ and $\mathbf{c}_j$, are ℓ_2-normalized. To compute pseudo-labels online, we apply the Sinkhorn-Knopp algorithm [10] to solve an entropy-regularized OT problem over the queue. Using the labels, we optimize the prototype loss:

$$\mathcal{L}_{\text{prototype}} = -\sum_{i=1}^{N} \sum_{j=1}^{M} Q_{ij} \log P_{ij}, \tag{3}$$

where Q_{ij} is taken for the current batch. This objective trains prototypes by aligning the model predictions to balanced assignments computed from the global queue. While this reflects the global morphological context, restricting the loss to the current batch prevents trivial assignment solution [1, 6].

2.3 Patch Search via Momentum Prototypes

Using the learned global prototype set $\mathbf{C}$, we dynamically retrieve k relevant patch features. We define the sampling index set $\mathcal{S}$ by a query-based search, where prototypes from $\mathbf{C}$ act as queries to find the top-k most relevant patch instances. Each prototype defines a corresponding sampling set of the most significant patches based on their relevance scores calculated using cosine similarity.

2.4 Training MMPL

We incorporate both learning global prototypes and classification simultaneously. Given the slide-level label Y, the loss consists of the supervised log-loss for $\hat{Y}$, defined as $\mathcal{L}_{\text{CE}} = -\sum_i Y_i \log \hat{Y}_i$, and $\mathcal{L}_{\text{prototype}}$ from Equation 3. Namely, the final loss $\mathcal{L} = \mathcal{L}_{\text{CE}} + \lambda \mathcal{L}_{\text{prototype}}$, where we set $\lambda = 0.1$ for all experiments.

For comprehensive evaluation, MMPL is tested in two scenarios: (1) freezing ϕ while training only θ to ensure comparability with other MIL methods, and (2) enabling end-to-end online training to jointly optimize all parameters. For the second scenario, we first freeze ϕ and retrieve $\mathcal{S}$, and then recompute the selected features by unfreezing ϕ for the gradient update of vision encoder.

Table 1: Performance evaluation of the two-step approach.

Model	CAMELYON16		PANDA(K)	PANDA(R)	TCGA-NSCLC	
	WF1	**AUC**	**Kappa**	**Kappa**	**WF1**	**AUC**
Backbone: ResNet50						
ABMIL [13]	0.844 (±.008)	0.887 (±.013)	0.835 (±.005)	0.820 (±.015)	0.821 (±.006)	0.894 (±.002)
DSMIL [15]	0.812 (±.015)	0.848 (±.016)	0.754 (±.023)	0.791 (±.026)	0.837 (±.008)	0.925 (±.003)
ZoomMIL [23]	0.858 (±.008)	0.869 (±.009)	0.860 (±.008)	0.811 (±.018)	0.834 (±.010)	0.912 (±.007)
IBMIL(abmil) [17]	0.863 (±.009)	0.898 (±.006)	0.414 (±.204)	0.456 (±.214)	0.851 (±.010)	0.932 (±.002)
ACMIL [27]	0.835 (±.016)	0.859 (±.011)	0.864 (±.004)	0.830 (±.009)	0.845 (±.005)	0.931 (±.004)
PANTHER [22]	0.786 (±.033)	0.829 (±.037)	0.695 (±.021)	**0.887** (±.010)	0.830 (±.006)	0.926 (±.001)
AEM [28]	0.834 (±.009)	0.857 (±.007)	**0.867** (±.003)	0.854 (±.006)	0.830 (±.009)	0.914 (±.001)
MMPL (Ours)	**0.905** (±.013)	**0.916** (±.013)	0.842 (±.014)	0.841 (±.039)	**0.869** (±.008)	**0.946** (±.006)
Backbone: UNI						
PANTHER [22]	0.808 (±.014)	0.849 (±.006)	0.923	0.931	0.934 (±.001)	**0.982** (±.000)
MMPL (Ours)	**0.987** (±.005)	**0.999** (±.001)	**0.934** (±.006)	**0.948** (±.004)	**0.950** (±.030)	0.980 (±.003)

3 Experiments

3.1 Experimental Setup

Datasets. Following prior work [23], we process WSIs by extracting 256×256 patches at ×20 magnification. We evaluate our method on three popular WSI datasets. (1) **CAMELYON16** [3] is a widely used benchmark dataset from the Cancer Metastases in Lymph Nodes Challenge, aimed at detecting metastatic breast cancer. The dataset consists of 399 WSIs of lymph node biopsies, with annotated regions indicating cancer presence. (2) **PANDA** [5] is a large-scale prostate cancer grading dataset of 9,555 WSIs across six ISUP grades from two institutions (Karolinska Institute and Radboud University Medical Center), evaluated using Quadratic Weighted Kappa. (3) **TCGA-NSCLC** [24], a lung cancer subset of The Cancer Genome Atlas dedicated to lung cancer, includes 1,042 WSIs from two subtypes: 530 LUAD and 512 LUSC. We report results under 5-fold cross-validation for UNI and cross-validation with standalone test for ResNet50, where five fold-trained models are aggregated by majority voting.

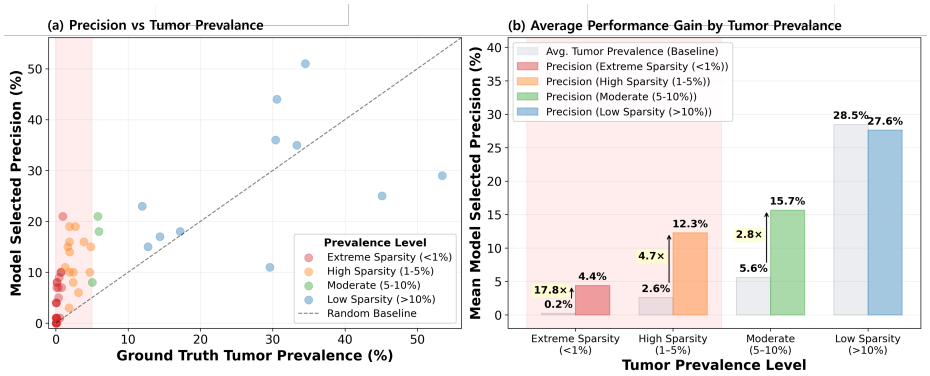

Fig. 2: (a) Each point denotes a single WSI, colored by tumor prevalence: extreme sparsity (red, $< 1\%$), high (orange, 1–5%), moderate (green, 5–10%), and low (blue, $> 10\%$). The dashed line ($y = x$) indicates the random baseline. (b) For each prevalence level, the left bar shows the average ground-truth tumor prevalence and the right bar shows the average precision of the model's top-k patches. Arrows denote the enrichment factor over ground truth.

Implementation Details. We trained using Adam with learning rates of 1×10^{-4} (MIL) and 1×10^{-5} (encoder), halved after 10 epochs without validation loss improvement. Models are trained for 100 epochs with best-checkpoint selection. The input, hidden, and attention dimensions are 1024, 512, and 256, respectively.

For prototype learning, we used 15 prototypes and retrieve the top 100 candidates. The queue size ranges from 100k to 200k depending on the dataset.

3.2 Results

We primarily adopt ResNet50 as the main encoder and additionally report results using UNI. Models are evaluated with ACC, WF1, and AUC. Table 1 summarizes the comparison across three datasets. With the ResNet50 backbone, our method achieves the best performance on CAMELYON16, improving over the baseline by +4.2%p in WF1 and +1.8%p in AUC. On TCGA-NSCLC, it also attains the highest performance with gains of +1.8%p in WF1 and +1.4%p in AUC. On PANDA, the proposed method remains competitive despite not reaching the top score. When using the UNI backbone, MMPL outperforms the prototype-based baseline on most metrics, with comparable performance on TCGA-NSCLC.

3.3 Further Analysis

Robustness to Low Tumor Prevalence. Figure 2 demonstrates robust performance under extremely low tumor prevalence on the CAMELYON16 test set. As shown in Figure 2(a), most slides, including those in the challenging 0–5% range, achieve precision above the random baseline. Quantitatively, Figure 2(b) reports enrichment factors of 2.8× to 17.8× for slides with less than 10%

Table 2: Distribution and annotated ratio of prototype assignments.

Distribution of Prototype Assignments														
P0	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**	**P10**	**P11**	**P12**	**P13**	**P14**
133k	167k	64k	123k	119k	74k	44k	116k	100k	116k	32k	83k	58k	37k	78k
Annotated Ratio per Prototype (%)														
P10	**P6**	**P13**	**P2**	**P7**	**P11**	**P9**	**P12**	**P4**	**P8**	**P3**	**P14**	**P0**	**P5**	**P1**
78.82	5.50	4.27	2.65	2.45	2.14	2.12	1.94	1.64	0.86	0.45	0.27	0.10	0.08	0.06

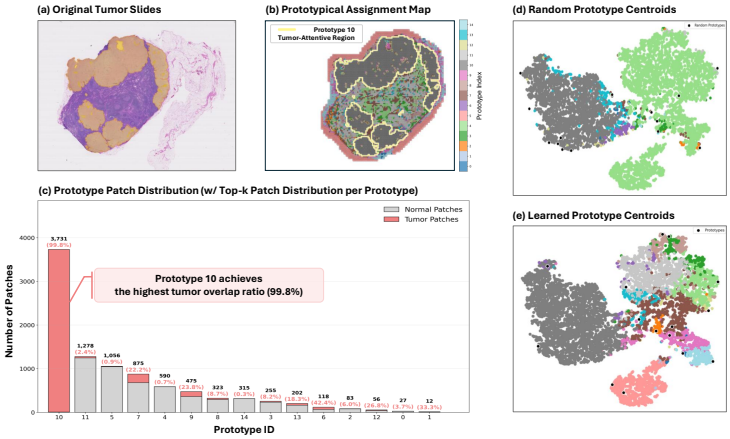

Fig. 3: Qualitative analysis of prototypical learning on CAMELYON16.

prevalence. These results confirm that the prototype-based retrieval effectively identifies rare tumor patches in highly imbalanced settings.

Semantic Interpretability of Prototypes. Table 2 presents the nearest-neighbor assignment distribution on the CAMELYON16 test set. Each prototype is assigned approximately 89.6k features on average, indicating balanced utilization without collapse. Prototype 10 shows the highest tumor-overlap ratio (78.82%), demonstrating effective clustering of tumor-relevant patches.

At the case level, Figure 3 visualizes prototype assignments for a representative tumor slide. Figure 3(a) shows the annotated tumor regions in yellow, and Figure 3(b) presents the corresponding prototype assignment map. Their spatial alignment indicates that the learned prototypes capture tumor morphology, including both core regions and boundaries. Consistent with global statistics (Table 2), Prototype 10 achieves the highest overlap ratio (99.8%) in this case (Figure 3(c)). The distribution of top-k retrieved patches, with the top-3 most similar patches per prototype illustrates the histological patterns across prototypes. Centroid visualization (Figure 3(d-e)) demonstrates a structured latent manifold with separated prototypes, unlike random initialization.

Table 3: Ablation study of queue, prototypes, and clustering. ResNet50 was used.

	CAMELYON16		
Model	**ACC**	**WF1**	**AUC**
w/o queue	0.745(±0.064)	0.743(±0.064)	0.764(±0.078)
w/o prototype	0.779(±0.036)	0.775(±0.036)	0.812(±0.030)
uniform top-k	0.846(±0.057)	0.845(±0.058)	0.864(±0.056)
K-means	0.812(±0.031)	0.812(±0.032)	0.865(±0.016)
MMPL (Ours)	**0.905(±0.013)**	**0.905(±0.013)**	**0.916(±0.013)**

Table 4: End-to-end comparison of VIB and MMPL. ResNet50 was used.

	CAMELYON16		
Model	**ACC**	**WF1**	**AUC**
VIB [16]	0.870(±0.009)	0.869(±0.010)	0.928(±0.013)
MMPL (Ours)	**0.916(±0.014)**	**0.916(±0.014)**	**0.976(±0.009)**

Ablation Study. As shown in Table 3, removing the feature queue degrades performance, confirming its importance for diverse prototype learning and generalization. Among prototype variants, the prototype-free model performs worst, validating the necessity of clustering, and uniform top-k underperforms dynamic top-k, demonstrating the benefit of selective prototype matching. Replacing Sinkhorn-Knopp–based online clustering with K-means further reduces performance, highlighting the advantage of dynamic optimal transport.

Comparison with Prior End-to-End Methods. While VIB [16] performs end-to-end fine-tuning, its decoupled pipeline prevents joint optimization, as it learns an Information Bottleneck mask on a frozen encoder before sampling. As shown in Table 4, our MMPL outperforms VIB by coupling patch selection, prototype learning, and slide-level prediction within one objective, enabling end-to-end optimization of the vision encoder.

Computational Complexity. To evaluate the training efficiency, we compare the peak memory and computational complexity (in TFLOPs) during the backward pass against the standard ABMIL baseline. Using a ResNet50 backbone with $N = 10,000$ patches, standard ABMIL requires storing full activations for all patches, demanding an impractical peak memory of 1.06 TB and 157.11 TFLOPs. In contrast, MMPL restricts encoder gradient updates to the top-k ($k = 100$) patches. Consequently, the peak memory drops to 22.26 GB, and the computational complexity decreases to 55.28 TFLOPs. This reduction enables end-to-end optimization on a single NVIDIA RTX 3090 GPU under this setting.

4 Conclusion

In this work, we presented MMPL, a prototype-based clustering framework for explicit patch selection in WSIs. By assigning patch features to global morphological prototypes via optimal transport and aggregating the top-k patches, MMPL enables structured, discriminative slide representation. Experiments show consistent improvements over baselines in metastasis detection, tumor grading, and subtyping, while supporting end-to-end training. Furthermore, it exhibits robustness under extremely low tumor prevalence and aligns prototypes with pathological regions without collapse. Extending this framework to tasks such as survival analysis and mutation prediction remains important future work.

Acknowledgments. This research was supported by (1) the National Research Foundation of Korea (NRF2023R1A2C3004176), (2) the Ministry of Health & Welfare, Republic of Korea (HR20C002103), (3) ICT Creative Consilience Program through the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (IITP-2026-RS2020-II201819), (4) the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT and MOE) (No. RS-2025-16652968), (5) the Seoul National University Hospital with support from the Ministry of Science and ICT (RS-2023-00262002) with computing resources including technical support.

Disclosure of Interests. The authors have no competing interests.

References

1. Asano, Y.M., Rupprecht, C., Vedaldi, A.: Self-labelling via simultaneous clustering and representation learning. In: International Conference on Learning Representations (2020)
2. Beck, A.H., Sangoi, A.R., Leung, S., Marinelli, R.J., Nielsen, T.O., Van De Vijver, M.J., West, R.B., Van De Rijn, M., Koller, D.: Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. Science translational medicine **3**(108), 108ra113–108ra113 (2011)
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. Bera, K., Schalper, K.A., Rimm, D.L., Velcheti, V., Madabhushi, A.: Artificial intelligence in digital pathology—new tools for diagnosis and precision oncology. Nature reviews Clinical oncology **16**(11), 703–715 (2019)
5. Bulten, W., Kartasalo, K., Chen, P.H.C., Ström, P., Pinckaers, H., Nagpal, K., Cai, Y., Steiner, D.F., Van Boven, H., Vink, R., et al.: Artificial intelligence for diagnosis and gleason grading of prostate cancer: the panda challenge. Nature medicine **28**(1), 154–163 (2022)
6. Caron, M., Misra, I., Mairal, J., Goyal, P., Bojanowski, P., Joulin, A.: Unsupervised learning of visual features by contrasting cluster assignments. Advances in neural information processing systems **33**, 9912–9924 (2020)

7. Caron, M., Touvron, H., Misra, I., Jegou, H., Mairal, J., Bojanowski, P., Joulin, A.: Emerging Properties in Self-Supervised Vision Transformers . In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 9630–9640. IEEE Computer Society, Los Alamitos, CA, USA (Oct 2021)
8. Chen, R.J., Ding, T., Lu, M.Y., Williamson, D.F., Jaume, G., Chen, B., Zhang, A., Shao, D., Song, A.H., Shaban, M., et al.: Towards a general-purpose foundation model for computational pathology. Nature Medicine (2024)
9. Chen, X., Fan, H., Girshick, R., He, K.: Improved baselines with momentum contrastive learning (2020), https://arxiv.org/abs/2003.04297
10. Cuturi, M.: Sinkhorn distances: Lightspeed computation of optimal transport. In: Burges, C., Bottou, L., Welling, M., Ghahramani, Z., Weinberger, K. (eds.) Advances in Neural Information Processing Systems. vol. 26. Curran Associates, Inc. (2013)
11. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum contrast for unsupervised visual representation learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9729–9738 (2020)
12. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
13. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International conference on machine learning. pp. 2127–2136. PMLR (2018)
14. Lerousseau, M., Vakalopoulou, M., Deutsch, E., Paragios, N.: Sparseconvmil: sparse convolutional context-aware multiple instance learning for whole slide image classification. In: MICCAI Workshop on Computational Pathology. pp. 129–139. PMLR (2021)
15. 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)
16. Li, H., Zhu, C., Zhang, Y., Sun, Y., Shui, Z., Kuang, W., Zheng, S., Yang, L.: Task-specific fine-tuning via variational information bottleneck for weakly-supervised pathology whole slide image classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7454–7463 (2023)
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. Niazi, M.K.K., Parwani, A.V., Gurcan, M.N.: Digital pathology and artificial intelligence. The lancet oncology **20**(5), e253–e261 (2019)
20. Rymarczyk, D., Pardyl, A., Kraus, J., Kaczyńska, A., Skomorowski, M., Zieliński, B.: Protomil: Multiple instance learning with prototypical parts for whole-slide image classification. In: Joint European conference on machine learning and knowledge discovery in databases. pp. 421–436. Springer (2022)
21. 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. In: Advances in Neural Information Processing Systems. vol. 34, pp. 2136–2147. Curran Associates, Inc. (2021)

22. Song, A.H., Chen, R.J., Ding, T., Williamson, D.F., Jaume, G., Mahmood, F.: Morphological prototyping for unsupervised slide representation learning in computational pathology. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11566–11578 (2024)
23. Thandiackal, K., Chen, B., Pati, P., Jaume, G., Williamson, D.F., Gabrani, M., Goksel, O.: Differentiable zooming for multiple instance learning on whole-slide images. In: European Conference on Computer Vision. pp. 699–715. Springer (2022)
24. Tomczak, K., Czerwińska, P., Wiznerowicz, M.: The cancer genome atlas (tcga): an immeasurable source of knowledge. Contemporary Oncology (Poznan) **19**(1A), A68–A77 (2015)
25. Yamamoto, Y., Tsuzuki, T., Akatsuka, J., Ueki, M., Morikawa, H., Numata, Y., Takahara, T., Tsuyuki, T., Tsutsumi, K., Nakazawa, R., et al.: Automated acquisition of explainable knowledge from unannotated histopathology images. Nature communications **10**(1), 5642 (2019)
26. Yang, L., Mehta, D., Liu, S., Mahapatra, D., Ieva, A.D., Ge, Z.: Trainable prototype enhanced multiple instance learning for whole slide image classification. In: Medical Imaging with Deep Learning. vol. 227, pp. 1655–1665. PMLR (2024)
27. 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)
28. Zhang, Y., Li, H., Sun, Y., Shui, Z., Li, J., Zhu, C., Yang, L.: AEM: Attention Entropy Maximization for Multiple Instance Learning based Whole Slide Image Classification . In: proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2025. vol. LNCS 15966. Springer Nature Switzerland (September 2025)
29. Zhu, X., Yao, J., Huang, J.: Deep convolutional neural network for survival analysis with pathological images. In: 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 544–547 (2016)