CLEAR-WSI: Towards Foundation-Model-Empowered Diagnosis Aligned Whole Slide Image Retrieval

Youssef Wally[‡1,2] ⋆, Jingsong Liu[‡1,3,6], Han Li[1,3,4], Jing Dai[5], Elisabeth Wetzer[2], and Peter J. Schüffler[1,3,6]

[1] Institute of Pathology, Technical University of Munich, TUM School of Medicine and Health, Munich, Germany
[2] Department of Physics and Technology, UiT The Arctic University of Norway, Tromsø, Norway
[3] Munich Center for Machine Learning (MCML), Munich, Germany
[4] Computer Aided Medical Procedures (CAMP), Technical University of Munich, Munich, Germany
[5] School of Biomedical Engineering, Faculty of Medicine, Dalian University of Technology, Dalian, China
[6] Munich Data Science Institute (MDSI), Munich, Germany

Abstract. The rapid growth of digital pathology has produced vast repositories of hematoxylin and eosin (H&E) stained whole slide images. However, many of these archives lack structured indexing and reliable metadata, making systematic organization and retrieval challenging. Reverse image search, also known as content-based image retrieval, addresses this limitation by retrieving slides based on learned visual representations rather than incomplete or inconsistent metadata. Although retrieval systems have been deployed in digital pathology, many rely on manually engineered indexing strategies or heuristic filtering rules, which limit scalability and generalization across diverse diagnostic categories. Thus, we propose CLEAR-WSI (Constant Length Embedding & Automatic Retrieval), a fully automated slide-level retrieval framework. CLEAR-WSI learns semantically structured whole slide embeddings through an Attention-based Multiple Instance Learning framework, compressing each WSI into a fixed dimensional representation for efficient storage and scalable similarity search. Furthermore, we introduce a self-reviewing diagnostic filtering mechanism that enforces label consistency among retrieved candidates, improving diagnosis alignment without relying on manually defined class-specific rules. Across two public datasets, CAMELYON16 (lymph node metastases) and BRACS (breast cancer subtypes), our diagnostic-aware method establishes new state-of-the-art results, improving Acc_{MV}@5 from 77.49% to 89.92% on CAMELYON16, from 54.12% to 75.86% on BRACS level-1, and from 36.47% to 51.72% on BRACS level-2. Our annotation-free, dataset-agnostic search engine that scales across diverse data sources is openly available: github.com/youssefwally/CLEAR-WSI

⋆ Corresponding author
[1] ‡These authors contributed equally to this work.

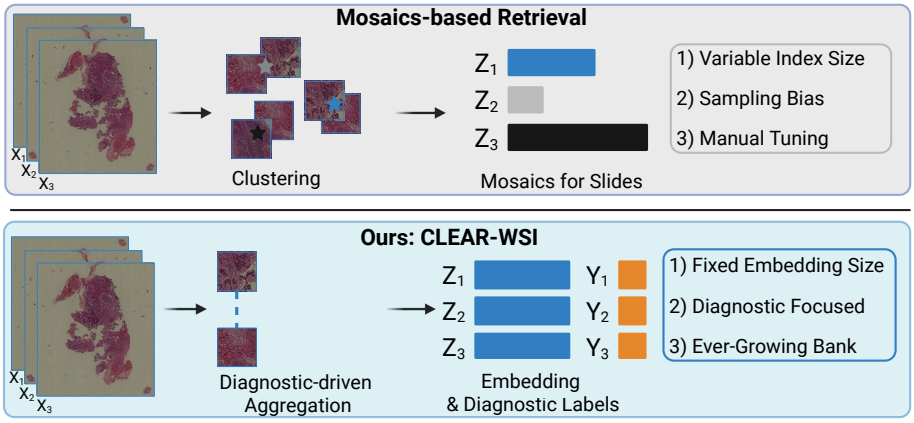

Fig. 1. Conceptual comparison between traditional mosaic-based methods and CLEAR-WSI. (Top) Conventional retrieval relies on manual clustering and sampling, leading to variable index sizes (Z_i) and sampling bias. **(Bottom)** `CLEAR-WSI` employs a compact diagnostic-driven aggregation that maps global context into a **fixed embbeding size** (Z_i), enabling automated, diagnosis-aligned search.

Keywords: Image Retrieval · Whole Slide Images · Breast Cancer.

1 Introduction

The digitization of pathology workflows has led to the rapid accumulation of whole slide images (WSIs) in hospital archives and public repositories. Beyond diagnosis, these collections enable case-based reasoning and similarity-driven exploration [15,13]. However, retrieving clinically meaningful similar slides at the whole slide level remains challenging due to the gigapixel scale and heterogeneous tissue composition of WSIs.

Unlike natural images, a single WSI may contain thousands of diverse tissue regions, requiring effective aggregation into a coherent slide-level representation [14]. Early content-based image retrieval systems (CBIR) relied on handcrafted texture and color features, while more recent approaches employ deep convolutional embeddings, hashing strategies, or attention-based multiple-instance learning frameworks [4,22]. Although these methods improve local discriminability, they often produce patch-level or variable-length representations that complicate large-scale indexing and limit diagnosis-level alignment.

Dedicated WSI-level retrieval systems such as Yottixel [11] introduce clustering based mosaics and bag-of-words indexing to improve scalability. Nevertheless, such approaches depend on manually configured clustering parameters and sampling strategies, and their slide representations vary with patch selection. Consequently, slide-level retrieval remains constrained by representation inconsistency, storage inefficiency, and reliance on heuristic filtering rules.

Despite architectural differences, existing CBIR systems share common structural limitations that affect scalability, reproducibility, and diagnosis-level consistency in WSI retrieval [20,12]: (1) **Manual hyperparameters and sampling bias:** Methods that rely on clustering + sampling (mosaics) require manual choices (number of clusters, sampling fraction) that affect coverage and reproducibility. (2) **Inconsistent representation sizes:** Index sizes that scale with the number of selected patches or mosaics produce variable memory and compute requirements. (3) **Limited context:** Patch-level retrieval systems may miss slide-level context important for diagnostics due to simple aggregation; losing spatial or structural information. (4) **Scalability vs. accuracy:** Compact/hashed indices and binarization increase speed but can degrade fine-grained retrieval necessary for rare or subtle morphologies.

We address these limitations by introducing `CLEAR-WSI` (**C**onstant-**L**ength **E**mbedding & **A**utomatic **R**etrieval), a fully automated framework for whole slide search (see Fig. 1). Central to our approach is a Diagnostic-driven Aggregation (**DX-Agg**) mechanism, which leverages attention-based pooling to holistically integrate the entire slide's context into a consistent, fixed-length representation. Unlike mosaic-based methods, our aggregation strategy eliminates sampling bias and ensures a predictable memory footprint for large-scale databases. Furthermore, we propose a **Self-review (SR)** filter to stabilize retrieval according to diagnosis. Evaluation on two datasets demonstrates that `CLEAR-WSI` establishes a new state-of-the-art (SOTA) for WSI retrieval, significantly improving ranking quality as measured by Normalized Discounted Cumulative Gain (NDCG) [19]. In summary, our main contributions are: (1) **Fully Automated Pipeline:** We introduce `CLEAR-WSI`, a search engine that eliminates manual hyperparameter tuning and sampling bias. (2) **Compact Diagnostic Integration:** We propose a unified pipeline combining **DX-Agg** and **SR** filtering, which maps variable-count features into a constant-length embedding while stabilizing retrieval through diagnostic refinement. (3) **SOTA Performance:** We establish new benchmarks on two public datasets and release our implementation for community use.

2 Methodology

CLEAR-WSI is designed to identify and rank Whole Slide Images (WSI) or their patches based on similarity to a given image query as illustrated in Figure 2. The process consists of (1) hierarchical feature extraction and aggregation, (2) similarity-based ranking, (3) label consistency filtering, and (4) storage of the resulting outputs in a growing memory bank for future use. This modular design enables the framework to operate at both the slide and patch levels, depending on whether patch selection and aggregation are enabled.

Hierarchical Feature Extraction and Aggregation. Direct end-to-end processing of a WSI is computationally prohibitive [15]. To address this, we employ a multi-stage feature extraction pipeline that prioritizes diagnostically relevant regions while leveraging the representative power of foundation models (FMs).

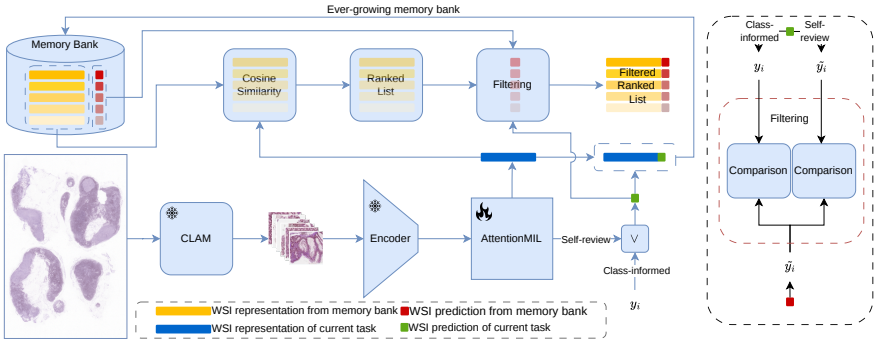

Fig. 2. Our pipeline, CLEAR-WSI, takes a WSI as a query, extracts relevant patches using CLAM [16] for which features are learned by a Foundation Model Encoder. An AttentionMIL classifier is trained on these features to produce a constant-length feature embedding relevant for a particular disease diagnosis, which serves as the representation for the retrieval database together with the label prediction (self-review) or ground truth labels if available (class-informed). Details in Section 2.

First, we perform tissue segmentation and tessellate the WSI into a set of patches $X_l = \{p_1, p_2, \ldots, p_N\}$ [16]. To ensure the final slide representation is both compact and semantically rich, we utilize a two-step embedding process. Unlike traditional methods that rely on class-conditioned embeddings, which often distort morphological similarity, we re-embed the most informative patches chosen by CLAM [16] using a foundation model (Foundation Models (FM)) encoder f_θ (e.g., UNI [5]). This produces a set of embeddings $\{z_j^{FM} = f_\theta(p_j)\}$, ensuring that the feature space remains uniform and independent of specific downstream labels.

To aggregate these local features into a global representation, we train an AttentionMIL module [9]. This module assigns a learnable attention weight β_j to each patch, capturing its contribution to the overall slide-level context. The final fixed-length global representation Z_l for each WSI is computed as:

$$Z_l = \sum_{j=1}^{|X_l|} \beta_j z_j^{FM} \tag{1}$$

By integrating FM encoders with AttentionMIL aggregation, `CLEAR-WSI` produces morphology-aware embeddings that effectively handle intra-slide heterogeneity while suppressing non-informative background noise.

Ranking by Similarity. Given a query embedding Z_q and a database $\mathcal{D} = \{Z_1, \ldots, Z_N\}$, similarity is measured using cosine similarity (CS). The ranking is defined as:

$$\pi(q) = \text{argsort}_{i=1..N}\ \text{CS}(Z_q, Z_i). \tag{2}$$

Diagnostic-aware Filtering. To enhance retrieval precision, we apply label-consistency strategies to refine the candidate list. The selection depends on whether the query slide label is available at inference time:

(1) Consistency-Inference (CI) Filtering: If the query label y_q is known (e.g., in expert-led verification), we enforce consistency using the ground truth:

$$\pi_{\mathrm{CI}}(q) = \{i \in \pi(q) \mid \tilde{y}_i = y_q\}. \tag{3}$$

(2) Self-review (SR) Filtering: In fully automated scenarios where the query label is unknown, we utilize the pseudo-label $\tilde{y}_q$ predicted by the AttentionMIL classifier:

$$\pi_{\mathrm{SR}}(q) = \{i \in \pi(q) \mid \tilde{y}_i = \tilde{y}_q\}. \tag{4}$$

Lastly, the top-k items from $\tilde{\pi}(q)$ form the retrieval candidate set $\tilde{\pi}_k(q)$, where $\tilde{\pi}(q)$ is either $\pi(q)$, $\pi_{SR}(q)$, $\pi_{CI}(q)$ depending on the filtering method.

Dynamic Memory Bank. To store embeddings and predicted pseudo-labels for efficient retrieval, a dynamic and ever-growing memory bank is constructed. This component transforms the framework from a fixed retrieval system into a dynamic knowledge base through an incremental update mechanism:

(1) Incremental Storage: Each query processed through the retrieval pipeline is not merely compared against existing entries; it is subsequently integrated back into the memory bank. The system stores the query's aggregated WSI embedding Z_q alongside its predicted pseudo-label $\tilde{y}_q$, allowing the repository to expand continuously as more cases are analyzed.

(2) Knowledge Evolution: Over time, this mechanism enables the system to benefit from a richer and more diverse set of stored representations. Such expansion enhances the framework's adaptability to evolving data distributions and inter-institutional variability, effectively refining the "diagnosis alignment" by increasing the density and diversity of the reference latent space.

3 Experimental Setup

3.1 Datasets

This study leverages two publicly available Hematoxylin and Eosin (H&E)-stained datasets widely used in contemporary computational pathology, CAMELYON16 dataset [1] and BReAst Carcinoma Subtyping (BRACS) dataset [3]. The datasets differ in their clinical diagnosis, scale, and labelling schemes, thus providing complementary challenges for evaluating our framework. Dataset splits follow the official releases.

3.2 CLEAR-WSI Setup

Encoder Selection. Four different ViT-based FM are ultilized in this stage; DeiT [21], MoCov3 [6], Prov-GigaPath [23] and UNI [5]. DeiT and MoCov3 are trained on general images, with DeiT using a distillation-based objective and MoCov3 using a self-supervised architecture on ImageNet-1K, providing efficient general-purpose visual representations but limited transfer to histopathology. Thus, we finetune both on CAMELYON16 [1] and BRACS [3]. In contrast, Prov-GigaPath and UNI are pathology FMs. Prov-GigaPath is trained on $\sim$1.3 billion

tiles from more than 170,000 WSI. UNI is pretrained on more than 100 million tissue patches from over 100,000 diagnostic WSI. Therefore, we do not finetune them. These models differ primarily in domain specificity and scale: DeiT and MoCov3 capture general visual features, whereas Prov-GigaPath and UNI provide domain-specialized representations optimized for computational pathology.
Training. An AttentionMIL model is trained to aggregate patch-level embeddings into slide-level representations. Once trained, the model is applied to the validation set to construct a *memory bank*. For each slide, we store its embedding and its corresponding label (either predicted by the AttentionMIL or ground-truth, depending on the filtering mechanism) to serve as a retrieval database.
Filtering and Memory Bank. During inference, test slides are processed through the pipeline, and retrieval results are refined based on two label consistency strategies **Consistency-Inference (CI)** and **Self-review (SR)**. In both cases, memory bank entries utilize predicted pseudo-labels. Ground-truth labels for the test set are strictly reserved for final performance evaluation.

3.3 Evaluation Metrics

We assess retrieval performance by evaluating how effectively the pipeline ranks relevant WSIs. A WSI is considered relevant if its label matches the query image label. **(1) Majority-Vote Accuracy:** We report majority-vote accuracy ($Acc_{MV}@k$) with $k = 1, 3, 5$. **(2) Normalized Discounted Cumulative Gain (NDCG):** Though NDCG [10] is widely adopted as an evaluation metric in recommender system evaluation [17,2,24,7,18,8], it is rarely adopted in WSI CBIR evaluation. Unlike accuracy-based measures, NDCG explicitly accounts for the ranking position of relevant results, rewarding models that place correct matches earlier in the retrieval list. Thus, we follow [19] and report NDCG evaluation metrics.

4 Results

Superiority Over SOTA Methods. Across all evaluated datasets, `CLEAR-WSI` consistently outperforms existing SOTA frameworks, including Yottixel, SISH, and RetCCL (Table 1). On CAMELYON16, our self-reviewing (SR) configuration achieves 89.92% $\text{Acc}_{MV}@5$ and 89.07% NDCG@5. It improves $\text{Acc}_{MV}@5$ over SISH (85.01%), while SISH retains a slightly higher NDCG@5 (89.67% vs. 89.07%). The performance gap is most striking on the complex BRACS tasks: on BRACS-1, `CLEAR-WSI` improves $\text{Acc}_{MV}@5$ from the SOTA 54.12% (Yottixel) to 75.86%, while on the multi-class BRACS-2, it increases NDCG@5 from 33.53% to 43.19%. These gains suggest that attention-based aggregation combined with a filter mechanism captures superior diagnostic context compared to traditional sampling or hashing approaches.
Retrieval Stability Across Neighborhood Sizes (k). The performance advantage of `CLEAR-WSI` remains robust across ranks $k = 1, 3, 5$. At the most stringent rank ($k = 1$), our SR pipeline attains 89.92% accuracy and 90.02%

Table 1. Comparative Benchmarking: CLEAR-WSI vs. SOTA methods across three clinical tasks. Best results are in **bold**, second best are underlined.

	Methods	CAMELYON16 NDCG | Acc_{MV}	BRACS-1 NDCG | Acc_{MV}	BRACS-2 NDCG | Acc_{MV}
@k=1	Yottixel-K [11]	76.21|75.96	49.91|49.41	33.61|32.94
	SISH [4]	86.66|86.66	51.37|50.21	35.86|34.05
	RetCCL [22]	85.32|84.16	48.74|50.55	35.22|35.07
	CLEAR-WSI (SR UNI)	90.02|89.92	**67.00|66.67**	**44.24|43.68**
	CLEAR-WSI (CI UNI)	**96.93|96.90**	43.87|43.68	10.53|10.34
@k=3	Yottixel-K [11]	74.59|78.29	49.58|54.12	31.25|36.47
	SISH [4]	86.90|86.25	47.52|49.21	35.47|35.05
	RetCCL [22]	86.47|86.18	48.07|48.92	33.23|35.82
	CLEAR-WSI (SR UNI)	89.56|89.15	**66.82|72.41**	**44.94|48.28**
	CLEAR-WSI (CI UNI)	**97.69|99.22**	42.02|45.98	11.24|16.09
@k=5	Yottixel-K [11]	73.69|77.49	48.64|54.12	29.17|36.47
	SISH [4]	89.67|85.01	47.34|50.85	33.53|34.16
	RetCCL [22]	86.05|86.82	47.26|48.23	30.24|31.15
	CLEAR-WSI (SR UNI)	89.07|89.92	**66.98|75.86**	**43.19|51.72**
	CLEAR-WSI (CI UNI)	**97.23|98.93**	41.39|50.57	11.46|14.94

NDCG on CAMELYON16, maintaining a lead over SISH and RetCCL. Critically, our NDCG gains consistently parallel or exceed accuracy improvements, demonstrating that `CLEAR-WSI` not only identifies more correct matches but effectively prioritizes them at the top of the ranked list.

Filtering Dynamics: CI vs. SR. The class-informed (CI) filtering, which assumes access to query labels, yields near-ceiling performance on CAMELYON16 (98.93% Acc_{MV}@5), validating the quality of the learned embedding space. However, on the more heterogeneous BRACS datasets, CI does not consistently surpass SR. On BRACS-2, for instance, SR achieves 43.19% NDCG@5 compared to just 11.46% for CI. This suggests that the AttentionMIL pseudo-labels used in SR provide a more "morphologically consistent" constraint than ground-truth labels in tasks with high intra-class variability, making SR highly effective for fully automated deployment.

5 Ablation Study

We summarize the impact of encoder selection and the aggregation design on retrieval performance in Table 2. This evolution confirms that the combination of a diagnostic-aware aggregator and high-capacity foundation representations is essential for achieving high-fidelity, slide-level retrieval.

Impact of Feature Encoders. We compared several state-of-the-art backbones, including DeiT, MoCov3, GigaPath, and UNI. While GigaPath and MoCov3 show incremental improvements over the DeiT baseline, the UNI encoder

Table 2. Ablation study on CAMELYON16 showing the performance evolution of `CLEAR-WSI`. Results represent the average NDCG and Acc_{MV} across $k \in \{1, 3, 5\}$.

	Encoder				Aggregation		Metrics	
Pipeline	DeiT	MoCov3	GigaPath	UNI	PCA	AttentionMIL	NDCG	Acc_{MV}
DeiT + PCA	✓				✓		63.66	66.41
DeiT + Attn	✓					✓	66.86 (↑3.20)	69.00(↑2.59)
MoCov3 + Attn		✓				✓	67.06 (↑0.20)	70.54 (↑1.54)
GigaPath + Attn			✓			✓	69.04 (↑1.98)	70.03 (↓0.51)
CLEAR-WSI (ours)				✓		✓	89.55 (↑20.51)	89.66 (↑19.63)

provides a substantial performance leap, increasing NDCG by over 20% compared to the GigaPath-based pipeline. This suggests that the foundational representations captured by UNI are more effective for encoding the complex morphological features required for accurate WSI retrieval and diagnosis alignment.

Aggregation Mechanism. The transition from simple linear aggregation PCA to a diagnostic-driven AttentionMIL model consistently improves both retrieval and classification metrics. For instance, using AttentionMIL with the DeiT backbone improves NDCG from 63.66% to 66.86%. The synergy between the UNI encoder and the AttentionMIL aggregator underscores the importance of our diagnostic-driven approach, which effectively identifies informative regions within vast H&E archives and aligns them with clinical categories.

6 Conclusion and Discussion

Technical Novelties and Retrieval Advantages. Unlike existing methods that rely on fragmented steps of clustering and sampling, `CLEAR-WSI` is a **compact, end-to-end pipeline** that integrates feature aggregation and semantic filtering. Our framework offers two primary systemic advantages: (1) **Diagnostic-driven Aggregation:** By aggregating global context into a **constant-length embedding**, we eliminate the sampling bias and manual hyperparameter tuning inherent in mosaic-based systems. (2) **Synergistic Stability:** The synergy between our aggregation and SR filtering bridges the gap between retrieval speed and diagnostic accuracy, ensuring semantic coherence in large-scale archives. This integrated approach transforms WSI retrieval from a manual, biased process into a fully automated, scalable, and clinically-aligned solution.

Scalability and Clinical Deployment. The system's design offers significant practical flexibility. Unlike traditional approaches, our method eliminates manual tuning of cluster counts or sampling rates and remains invariant to patch count, simplifying deployment across diverse hardware and datasets. Benchmarking demonstrates high efficiency: on an NVIDIA GeForce RTX 3090, our pipeline processes a whole-slide image in $\sim$ 0.01s end-to-end, spanning WSI loading, CLAM patch selection, foundation-model encoding, and AttentionMIL aggregation; while patch retrieval takes < 0.01s. This confirms the framework's capacity to manage "ever-growing" memory banks with negligible overhead. While we uti-

lize a static validation bank here, the architecture inherently supports continual growth as new cases are processed.

Limitations and Future Work. Despite these strengths, SR filtering performance depends on a reasonably calibrated classifier. Future iterations could integrate uncertainty-aware filtering and out-of-distribution detection to enhance robustness against rare or novel morphologies. Additionally, while our initial results on patch-level queries are promising, prospective user studies are required to fully assess the interpretability and impact of this retrieval system on actual diagnostic workflows.

7 Acknowledgments

This work was supported by the German Federal Ministry of Education and Research (BMBF)-funded SATURN3 project (01KD2206C), the Innovative Medicines Initiative (IMI) BIGPICTURE project (IMI945358), the Research Council of Norway, Integreat – Norwegian Centre for knowledge-driven machine learning (project number 332645).

References

1. Bejnordi, B., Veta, M., Van Diest, P., Van Ginneken, B., Karssemeijer, N., Litjens, G., Van Der Laak, J., Hermsen, M., Manson, Q., Balkenhol, M., Geessink, O.: Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA **318**(22), 2199–2210 (2017)
2. Bellogín, A., Castells, P., Cantador, I.: Precision-oriented evaluation of recommender systems: An algorithmic comparison. In: Proceedings of the Fifth ACM Conference on Recommender Systems. pp. 333–336. ACM (2011)
3. Brancati, N., Anniciello, A., Pati, P., Riccio, D., Scognamiglio, G., Jaume, G., De Pietro, G., Di B., M., Foncubierta, A., Botti, G., Gabrani, M., Feroce, F., Frucci, M.: BRACS: A dataset for BReAst Carcinoma Subtyping in H&E histology images. Database **2022** (10 2022). https://doi.org/10.1093/database/baac093
4. Chen, C., Lu, M., Williamson, D.F., Chen, T., Schaumberg, A., Mahmood, F.: Fast and scalable search of whole-slide images via self-supervised deep learning. Nature Biomedical Engineering **6**(12), 1420–1434 (2022)
5. Chen, R., Ding, T., Lu, M., Williamson, D., Jaume, G., Chen, B., Zhang, A., Shao, D., Song, A., Shaban, M., et al.: Towards a general-purpose foundation model for computational pathology. Nature Medicine (2024)
6. Chen, X., Xie, S., He, K.: An empirical study of training self-supervised vision transformers. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (2021)
7. He, X., Deng, K., Wang, X., Li, Y., Zhang, Y., Wang, M.: LightGCN: Simplifying and powering graph convolution network for recommendation. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 639–648. ACM (2020)
8. He, X., Liao, L., Zhang, H., Nie, L., Hu, X., Chua, T.: Neural collaborative filtering. In: Proceedings of the 26th International Conference on World Wide Web. pp. 173–182. ACM (2017)
9. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International conference on machine learning. pp. 2127–2136. PMLR (2018)
10. Järvelin, K., Kekäläinen, J.: Cumulated gain-based evaluation of IR techniques. ACM Transactions on Information Systems (TOIS) **20**(4), 422–446 (2002)
11. Kalra, S., Tizhoosh, H., Choi, C., Shah, S., Diamandis, P., Campbell, C., Pantanowitz, L.: Yottixel–an image search engine for large archives of histopathology whole slide images. Medical Image Analysis **65**, 101757 (2020)
12. Lahr, I., Alfasly, S., Nejat, P., Khan, J., Kottom, L., Kumbhar, V., Alsaafin, A., Shafique, A., Hemati, S., Alabtah, G., Comfere, N., Murphree, D., Mangold, A., Yasir, S., Meroueh, C., Boardman, L., Shah, V., Garcia, J., Tizhoosh, H.: Analysis and validation of image search engines in histopathology. IEEE Reviews in Biomedical Engineering pp. 1–19 (2024). https://doi.org/10.1109/RBME.2024.3425769
13. Liu, J., Deng, X., Li, H., Kazemi, A., Grashei, C., Wilkens, G., You, X., Groll, T., Navab, N., Mogler, C., et al.: From pixels to pathology: Restoration diffusion for diagnostic-consistent virtual IHC. Computers in Biology and Medicine **198**, 111264 (2025)
14. Liu, J., Li, H., Xu, Z., Klaus, F.L., Stögbauer, F., Zu, S., Zhou, W., Kasajima, A., Schicktanz, F., Muckenhuber, A., Shakhtour, J., Yu, J., Zheng, T., Ma, X., Wang, M., Grashei, C., Li, B., Jiang, G., Xu, H., Zhou, S.K., Navab, N., Schüffler, P.J.: Towards cellular-scale interpretability in pathology foundation models for biomarker assessment (2025), https://arxiv.org/abs/2511.05150

15. Liu, J., Li, H., Yang, C., Deutges, M., Sadafi, A., You, X., Breininger, K., Navab, N., Schüffler, P.J.: HASD: hierarchical adaption for pathology slide-level domain-shift. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 332–342. Springer (2025)
16. Lu, M., Williamson, D., Chen, T., Chen, R., Barbieri, M., Mahmood, F.: Data-efficient and weakly supervised computational pathology on whole-slide images. Nature biomedical engineering **5**(6), 555–570 (2021)
17. Mao, J., Liu, X., He, X., Jin, D., Li, Y.: Simplex: A simple and strong baseline for collaborative filtering. In: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 103–112. ACM (2021)
18. Rendle, S., Freudenthaler, C., Gantner, Z., Schmidt-Thieme, L.: BPR: Bayesian personalized ranking from implicit feedback. In: Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence. pp. 452–461. AUAI Press (2009)
19. Shi, X., Sapkota, M., Xing, F., Liu, F., Cui, L., Yang, L.: Pairwise based deep ranking hashing for histopathology image classification and retrieval. Pattern Recognition **81**, 14–22 (2018)
20. Tizhoosh, H., Pantanowitz, L.: On image search in histopathology. Journal of Pathology Informatics **15**, 100375 (2024)
21. Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A., Jegou, H.: Training data-efficient image transformers & distillation through attention. In: International Conference on Machine Learning. vol. 139, pp. 10347–10357 (2021)
22. Wang, X., Du, Y., Yang, S., Zhang, J., Wang, M., Zhang, J., Yang, W., Huang, J., Han, X.: RetCCL: Clustering-guided contrastive learning for whole-slide image retrieval. Medical image analysis **83**, 102645 (2023)
23. Xu, H., Usuyama, N., Bagga, J., Zhang, S., Rao, R., Naumann, T., Wong, C., Gero, Z., González, J., Gu, Y., Xu, Y., Wei, M., Wang, W., Ma, S., Wei, F., Yang, J., Li, C., Gao, J., Rosemon, J., Bower, T., Lee, S., Weerasinghe, R., Wright, B., Robicsek, A., Piening, B., Bifulco, C., Wang, S., Poon, H.: A whole-slide foundation model for digital pathology from real-world data. Nature (2024)
24. Xue, H., Dai, X., Zhang, J., Huang, S., Chen, J.: Deep matrix factorization models for recommender systems. In: Proceedings of the 26th International Joint Conference on Artificial Intelligence. pp. 3203–3209. IJCAI (2017)