Meta-SSL: A Meta-Learning-Augmented Self-Supervised Framework for Histomorphological Phenotyping and WSI Retrieval

Panto Saha[1], Ishtiaque Ahmed[1]
[1]MSc in CSE-Data Science, Faculty of Science and Information Technology, DIU
{253-25-022@diu.edu.bd, 253-25-002@diu.edu.bd}

TL;DR: Meta-SSL proposes a lightweight four-stage framework that sequentially combines Barlow Twins self-supervised pretraining on 100K colorectal patches, first-order MAML meta-learning across mixed histological domains, and VQ-VAE discrete tokenization to achieve 96.52% kNN accuracy on CRC-VAL-HE-7K and +28–40% cross-domain few-shot gains over DINO, Lunit-DINO, and CTransPath, with sub-linear patch retrieval.

Problem and Objective

Clinical Problem

- Histopathological diagnosis demands expert visual inspection of Whole Slide Images (WSIs), which is labour-intensive, expensive, and prone to inter-observer variability.
- Large annotated datasets are scarce in computational pathology; most real-world scenarios involve only a handful of labelled examples per tissue class.

Technical Gap

- Existing SSL methods (SimCLR, DINO, CTransPath) learn strong features on large corpora but fail to adapt rapidly to unseen tissue domains or rare phenotypes with only 1–10 labelled shots.
- Standard kNN/ANN retrieval over continuous embeddings scales poorly in memory and query latency as WSI archives grow to millions of patches.

Objective

Design a lightweight, end-to-end trainable framework that learns rich domain-generalizable patch representations without labels and enables fast cross-domain adaptation via episodic meta-learning and also supports efficient discrete-code retrieval with sub-linear query complexity.

Methodology

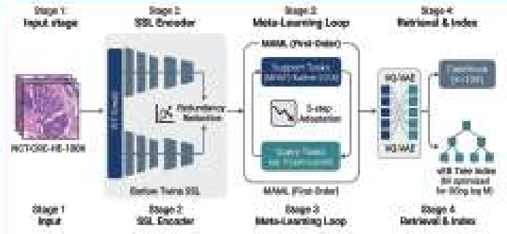

Fig. 1: The complete four-stage pipeline is illustrated, encompassing input preprocessing from the NCT-CRC-HE-100K dataset, self-supervised encoding via Barlow Twins, first-order MAML-based meta-learning, and VQ-VAE-driven retrieval indexed by van Emde Boas tree.

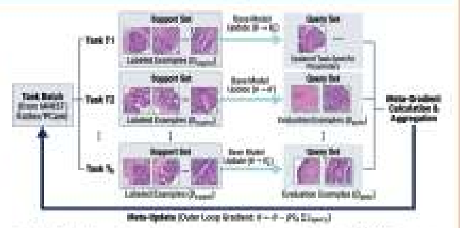

Fig. 3: The Stage 3 meta-learning loop adopts a first-order MAML strategy, iteratively adapting base model parameters across support and query task sets sampled from MHIST, and PCam datasets, with per-task gradients aggregated through an outer-loop meta-update rule.

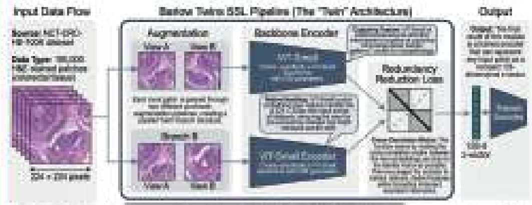

Fig. 2: The Stage 2 SSL encoder employs a dual-branch ViT-Small architecture trained under the Barlow Twins objective, minimizing feature redundancy through cross-correlation alignment to produce compact 128-dimensional z-vector representations of each 224×224 H&E-stained patch.

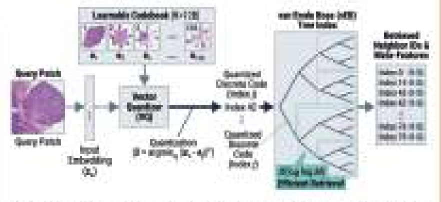

Fig. 4: The Stage 4 retrieval pipeline leverages a VQ-VAE equipped with a learnable codebook of K=128 prototypical embeddings to discretize patch representations into quantized discrete codes, enabling efficient nearest-neighbor retrieval via a van Emde Boas tree index.

Benchmark

Table 1: Incremental Stage Contribution

Stage	kNN acc (%) ↑	ARI ↑	NMI ↑	Silhouette ↑	Davies-Bouldin ↓
S1: SimCLR	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
S2: Barlow Twins (SSL)	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
S3: VQ-VAE [illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
S4: Meta-SSL (FOMAML)	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Table 2: Cross-domain few-shot accuracy

Method	1-shot (%)	5-shot (%)	10-shot (%)
DINO (ImageNet)	[illegible]	[illegible]	[illegible]
Lunit-DINO	[illegible]	[illegible]	[illegible]
CTransPath	[illegible]	[illegible]	[illegible]
SimCLR (S1)	[illegible]	[illegible]	[illegible]
Barlow Twins (S2)	[illegible]	[illegible]	[illegible]
Meta-SSL (S4)	[illegible]	[illegible]	[illegible]

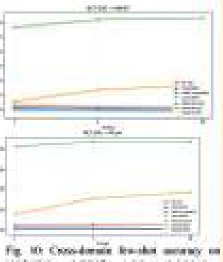

Fig. 10: Cross-domain few-shot accuracy on NCT-CRC → MHIST and PCam (1/5/10-shot, 600 trials)

Experiments and Results

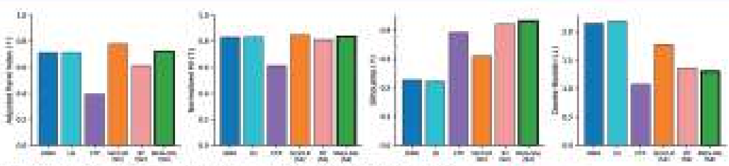

Fig. 5: Unsupervised clustering quality metrics across six methods on CRC-VAL-HE-7K

Table 3: Multi-Modal Retrieval performance

Method	Cosine ↑	VQ Code ↑	Fused ↑	Classes >50%
DINO (ImageNet)	94.2%	64.0%	88.1%	9/9
Lunit-DINO	94.3%	64.0%	88.1%	9/9
CTransPath	81.3%	64.0%	75.8%	9/9
SimCLR (S1)	92.3%	64.0%	88.0%	9/9
Barlow Twins (S2)	94.3%	64.0%	90.2%	9/9
Meta-SSL (S4)	**96.0%**	**64.0%**	**88.8%**	**9/9**

Fig. 6. VQ codebook utilization histogram showing 99.2% active codes (254/256) with EMA updates, confirming no codebook collapse.

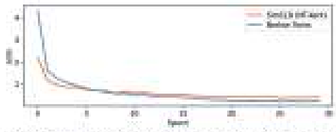

Fig. 8: Comparative SSL pretraining loss curves: SimCLR (NT-Xent) and Barlow Twins convergence over 30 epochs on NCT-CRC-HE-100K.

Supplementary Info

Fig. 7. Qualitative top-5 patch retrieval across nine tissue classes using fused cosine and VQ-code similarity. green = correct, red = incorrect.

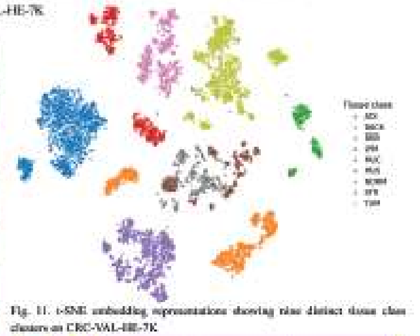

Fig. 11. t-SNE embedding representations showing nine distinct tissue class clusters on CRC-VAL-HE-7K

Fig. 12. VQ-VAE codebook usage by tissue class across 183 active codes, showing class-discriminative code specialization.

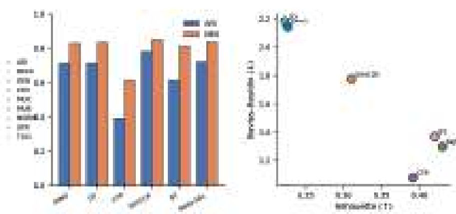

Fig. 9. Unsupervised alignment with ground truth (ARI/NMI) and cluster separation quality (Silhouette vs Davies-Bouldin) across all methods

REFEFENCES

[1] Chen, R.J., Ding, T., Lu, M.Y. et al. Towards a general-purpose foundation model for computational pathology. Nat Med 30, 850–862 (2024). https://doi.org/10.1038/s41591-024-02857-3

[2] Vorontsov, E., Bozkurt, A., Casson, A. et al. A foundation model for clinical-grade computational pathology and rare cancers detection. Nat Med 30, 2924–2935 (2024). https://doi.org/10.1038/s41591-024-03141-0

[3] Xu, H., Usuyama, N., Bagga, J. et al. A whole slide foundation model for digital pathology from real world data. Nature 630, 181–188 (2024). https://doi.org/10.1038/s41586-024-07441-w

ACKNOWLEDGEMENT

I sincerely thank my supervisor for guidance and support. I am grateful to the Department of SWE and HIRL Lab for providing resources and a research-friendly environment. Special thanks to DoR for the opportunity to present this paper.

Daffodil International University