Nine Dimensions Are Enough for Patch-Level Classification with Histopathology Foundation Model Embeddings

Mirza Nasir Hossain*,†, David Morrison*, Julia Dreiling, and David Harris-Birtill

School of Computer Science, University of St Andrews
mnh3@st-andrews.ac.uk

Abstract. Foundation Models (FMs) based on the self-supervised DINOv2 architecture have become the standard approach for feature extraction in computational pathology. In these models, embedding size is typically inherited from the FM and treated as a fixed design choice rather than a hyperparameter. However, there is growing evidence that embeddings produced by this class of FMs exist in a space with an intrinsic dimensionality substantially lower than the number of dimensions in the embedding. This suggests that embeddings can be projected into more compact representations, saving substantial computational resources for downstream tasks. This work evaluates three projection techniques (principal component analysis (PCA), uniform manifold approximation and projection (UMAP), and Gaussian random projection (GRP)) via downstream classification and utilises five datasets covering four tissue types (colorectal, breast, skin, thorax), and across four FMs: three histopathology FMs (UNI, UNI2-h, and Hibou-L) and DINOv2 as a general-purpose baseline. Across the 20 dataset–encoder pairs, PCA recovers 99% of full-dimensional classification ROC-AUC at a median target dimension of 9, reducing dense float32 feature storage by 113.8×–170.7× relative to 1024-D and 1536-D embeddings. Furthermore, compressing to the estimated intrinsic dimensionality preserves near-parity or surpasses full-dimensional performance for the histopathology FMs, whereas DINOv2 shows greater degradation at lower dimensions.

Keywords: Foundation Models · Dimensionality Reduction · Computational Pathology · PCA · Intrinsic Dimensionality.

1 Introduction

Foundation Models (FMs) based on the self-supervised DINOv2 [25] architecture have rapidly become the standard approach for feature extraction in computational pathology [29, 9]. Histopathology FMs such as UNI [4], UNI2-h [4] and

* These authors contributed equally.
† Corresponding author.

Hibou-L [22] extract high-dimensional embeddings from hematoxylin and eosin stained tissue patches (H&E), providing general-purpose representations for several downstream tasks, including cancer subtyping [12], tissue classification [28], and biomarker prediction [2]. More than twenty histopathology FMs are now publicly available [19], and they have demonstrated strong performance across diverse datasets [24], making them a widely adopted building block for computational pathology pipelines.

However, deploying pathology FMs at scale imposes practical compute and storage limits. Whole-slide images (WSIs) are gigapixel-scale and often several gigabytes at diagnostic resolution; they are tiled into 224×224-px patches, yielding $\sim$10k–40k patches per slide depending on magnification [17]. Each patch is encoded as a high-dimensional vector (typically 1024 D or 1536 D). With 1024-D float32 embeddings, one slide produces $\sim$40–160 MB of features; for a moderate cohort of 10,000 WSIs, this implies $\sim$0.4–1.6 TB of embedding storage per encoder. During downstream training, transferring these features between CPU and GPU memory becomes a bottleneck, and model memory and compute scale linearly with embedding dimensionality.

Despite these constraints, embedding dimensionality is typically inherited from the FM and treated as a fixed design choice rather than a hyperparameter [24]. While pruning, distillation, and quantization can reduce FM parameter counts and inference costs which typically requires retraining or fine-tuning the encoder, there's been little investigation into whether patch embeddings from frozen pretrained FMs can be projected to a lower-dimensional subspace without degrading performance. It remains unclear what the effective dimensionality of these representations is, whether this holds across tissues, and FM scale, and how different compression methods trade off accuracy, robustness, and efficiency.

Our contribution is empirical rather than algorithmic. We systematically evaluate dimensionality reduction for histopathology FM embeddings across four encoders, three projection methods, and five datasets spanning four tissue types. PCA recovers 99% of full-dimensional classification ROC-AUC at a median of just 9 dimensions, corresponding to 113.8×–170.7× lower dense feature storage for 1024-D and 1536-D embeddings. Histopathology FMs retain near-parity with full-dimensional baselines when compressed to their estimated intrinsic dimensionality, while DINOv2 shows larger degradation at low dimensionality.

2 Related Work

In the standard computational pathology pipeline, a pretrained encoder is frozen and its patch-level embeddings are passed to a lightweight downstream head that is trained for a target task. Models such as UNI, UNI2-h, and Hibou-L produce embeddings of 1024 or 1536 dimensions, a choice inherited from the ViT [7] architecture rather than derived from the input data. Recent benchmarking studies indicate that these models are converging on downstream task performance [2, 24], suggesting that further gains may not come from scaling encoders but from examining whether the resulting embeddings are efficiently structured.

For general image classification, Chandranegara et al. [3] evaluated dimensionality reduction of 384-D DINO-ViT features to as few as 5 dimensions using principal component analysis (PCA) [6], neighbourhood components analysis (NCA) [10], and uniform manifold approximation and projection (UMAP) [20]. In contrast, we evaluate modern histopathology FMs across multiple tissues and encoders. Retrieval-focused work has also evaluated pretrained CNN and foundation-model features for content-based medical image retrieval [18], but not AUC-preserving compression of histopathology FM embeddings across encoders and tissues. Modern histopathology FMs have also been shown to encode site-specific artifacts such as staining procedures and scanner differences [11], suggesting that not all embedding dimensions encode diagnostically relevant information.

Adjacent work has addressed the computational cost of high-dimensional embeddings through distillation [8] and vector quantization of spatial token sequences [16], but none target compression of the embeddings produced by the FM. Notably, Filiot et al. [8] observe explicitly that the intrinsic dimension of foundation model representations requires further investigation.

Experimental evidence suggests that deep neural network representations sit on a lower-dimensional manifold than their embedding dimensionality. Intrinsic dimensionality in trained networks is orders of magnitude smaller than layer width, particularly in final layers as shown by Ansuini et al. [1]. Estimates across several histopathology FMs are also consistent with substantial redundancy [27]. These findings suggest that histopathology FM embeddings can be projected to substantially lower dimensions while preserving downstream task performance, yet no evaluation of post-hoc dimensionality reduction has been conducted across multiple encoders, tissue types, and projection methods.

3 Methodology

The study consists of three stages: embedding, projection, and evaluation. Let I_i be an image patch and let $f_m(I_i) = \mathbf{x}_i \in \mathbb{R}^{d_m}$ be its embedding under encoder m. For projector p and target dimension k, a transform $T_{m,p,k} : \mathbb{R}^{d_m} \to \mathbb{R}^k$ was fitted using training embeddings only, giving $\mathbf{z}_i^{(k)} = T_{m,p,k}(\mathbf{x}_i)$. A downstream Logistic Regression (LR) classifier $h_{m,p,k}$ was then fitted on the projected training representations and evaluated on the unchanged held-out test split. The four encoders were UNI, UNI2-h, Hibou-L, and DINOv2; the projection methods were PCA, UMAP, and Gaussian Random Projection (GRP).

Default library settings were used unless noted [26]. GPU logistic regression frequently failed to converge on UMAP projections at certain target dimensionalities, so we performed a per-dimension training-only sweep on Kather100K embeddings (UNI2-h), testing inverse regularisation strength, solver tolerance, and max iterations near library defaults. The selected configuration was fixed and applied across all subsequent experiments; no test set was used for pro-

Table 1. Dataset summary. Counts refer to labelled central patches used for feature extraction; SPIDER also provides surrounding context patches, which were not used in this study.

Dataset	Organ	Classes	Train	Test	Total
Kather100K	Colorectal	9	100,000	7,180	107,180
SPIDER-CRC	Colorectal	13	63,989	13,193	77,182
SPIDER-Breast	Breast	18	80,858	12,034	92,892
SPIDER-Skin	Skin	24	131,164	28,690	159,854
SPIDER-Thorax	Thorax	14	63,319	14,988	78,307

jector or classifier hyperparameter selection. Code is available online[1]. Cached embeddings will be linked there when released.

Datasets

Five open-access datasets of labelled histopathology patches were selected for this study: Kather100k [13] and the skin, colorectal, thorax, and breast datasets from Supervised Pathology Image-DEscription Repository (SPIDER) [23]. All patch datasets are 224×224 pixel images extracted at 0.5 microns per pixel (MPP) from Whole Slide Images (WSIs).

Table 1 summarises the datasets. For SPIDER, we used the central labelled patches and the official train-test split, which is performed at slide level to prevent patches from the same WSI appearing in both splits.

Embedding

Vision FMs map each image patch to a fixed-length embedding vector. Dimensionality reduction methods then project these high-dimensional embeddings into a lower-dimensional space while aiming to preserve the structure needed for downstream analysis.

Image patches from each dataset were encoded using UNI [4] (widely adopted), UNI2-h [4] (state-of-the-art), and Hibou-L [22] (largest histopathology pretraining corpus), with DINOv2-L [25] as a general-purpose reference. DINOv2-L was included to compare pathology-specific and general-purpose self-supervised ViT encoders. This comparison does not isolate pretraining domain causally because the models also differ in training data, objectives, and scale. For each embedded dataset, the intrinsic dimensionality (ID) was estimated using the Maximum Likelihood Estimation (MLE) [15].

Projections

Dimensionality reduction was performed using three projection techniques: Principal Component Analysis (PCA), Uniform Manifold Approximation and Pro-

[1] https://github.com/davemor/hdims

jection (UMAP), and Gaussian Random Projection (GRP). PCA identifies orthogonal directions of maximal variance, UMAP preserves local neighbourhood structure as a nonlinear manifold method, and GRP employs a randomly generated projection matrix as a stochastic baseline.

For each method, embeddings were projected to multiple target dimensionalities from $k = 1$ up to $k = d$ rounded to the nearest fiftieth dimension, where d is the dimensionality of the original embedding space. We evaluated all dimensions for $k \leq 50$, and then at increments of 50. For every target dimensionality, the projection transform was learned using the training embeddings only.

For each dataset–encoder–projector combination, ROC-AUC retention and the recovery dimension are

$$R_k = \frac{\mathrm{AUC}_k}{\mathrm{AUC}_{\mathrm{full}}}, \qquad d_{99} = \min\{k : R_k \geq 0.99\}.$$

Thus, d_{99} is a descriptive operating point based on observed test performance, not a statistically established non-inferiority boundary.

UMAP requires hyperparameters that can materially affect downstream performance, so we used a training-only two-phase procedure: sweep a small set of target dimensions and hyperparameters, then re-run selected settings across the full dimensionality range. GRP results use one random projection per dataset–encoder–dimension and are not variance-characterised over projection seeds.

Evaluation

Downstream utility of the projections was evaluated via patch-level classification. We trained and evaluated an LR classifier and computed accuracy and ROC-AUC, with paired bootstrap confidence intervals for ROC-AUC (1000 resamples; 95% CI). We report our main analysis in terms of ROC-AUC because it is insensitive to class imbalance within the dataset.

4 Results

At full dimensionality (d=1536 for UNI2-h, d=1024 for all other FMs), Logistic Regression achieved a mean ROC-AUC of 0.969 with a standard deviation of 0.051 and a mean accuracy of 0.800 with a standard deviation of 0.163 across all patch-level datasets and encoders. Mean values were computed by evaluating each dataset–encoder pair separately and then averaging the resulting scores; datasets were not pooled before metric computation. The intrinsic dimensionality (ID) of the full embeddings, estimated using MLE, ranged from 9 (Hibou-L on Kather100K) to 20 (UNI2-h on SPIDER-CRC), with histopathology FMs averaging 14.6 and DINOv2 averaging 17.05, computed from the unrounded MLE estimates.

Because feature storage, memory footprint, and tensor-transfer volume scale linearly with embedding dimensionality, the practical storage impact can be estimated directly as $N \cdot k \cdot d_{\mathrm{bytes}}$, where N is the number of patches, k is the

Table 2. Analytical dense float32 feature-storage estimates. Ratio columns give reductions relative to full 1024-D or 1536-D embeddings and proportional reductions in feature-memory and CPU–GPU transfer volume; they are not wall-clock measurements.

Representation	Dim.	Bytes/patch	MB/10k patches	1024-D ratio	1536-D ratio
Full 1024-D	1024	4096	40.96	1.0×	–
Full 1536-D	1536	6144	61.44	–	1.0×
PCA median d_{99}	9	36	0.36	113.8×	170.7×
PCA $k = 50$	50	200	2.00	20.5×	30.7×
Best observed d_{99}	4	16	0.16	–	384.0×

embedding dimension, and $d_{\text{bytes}} = 4$ for float32 features. Table 2 reports analytical dense-array estimates, excluding file-format metadata, indexing, compression, and filesystem overhead. The 384× ceiling corresponds to the best observed case in Table 3: UNI2-h has 1536-D embeddings and recovers 99% ROC-AUC at $d_{99} = 4$ on Kather100K. The median $d_{99} = 9$ corresponds to 113.8×–170.7× smaller dense feature arrays depending on the encoder.

As embeddings are projected into lower dimensions, aggregate classification ROC-AUC remains consistent above 50 dimensions. Across the 20 patch-level dataset–encoder pairs, PCA retains 99% of full-dimensional ROC-AUC at a median target dimension of 9. Figure 1 shows the mean Logistic Regression ROC-AUC across all foundation models and datasets at each dimension for each projector. At the aggregate level, mean performance approaches 99% retention at 50 dimensions for all three projectors; this does not imply that every pair reaches the threshold.

Table 3 shows an expanded view of the minimum dimensions for 99% of their full dimensional ROC-AUC retention for each dataset shown for PCA, UMAP, and GRP. Histopathology FMs and DINOv2 differ markedly in their compression behaviour. For example, on SPIDER-Breast, histopathology FMs recover around 8–10 dimensions with PCA while DINOv2 requires 100. Under UMAP, DINOv2 reaches the 99% threshold only on Kather100K.

Figure 2 summarises performance at 9 dimensions, the median dimension that recovers 99% of the full dimensional ROC-AUC. PCA at $d = 9$ achieves near-parity with full-dimensional baselines for histopathology FMs across datasets, while DINOv2 exhibits a consistent performance gap. On Kather100K, representative paired-bootstrap 95% CIs for full versus $d = 9$ ROC-AUC were 0.994–0.996 versus 0.996–0.997 for UNI, and 0.990–0.993 versus 0.979–0.982 for DINOv2. These marginal intervals are not a paired non-inferiority test.

To test whether this behaviour depended on the downstream classifier, we repeated the PCA $d = 9$ comparison using Multi-layer Perceptron (MLP) and XGBoost [5] classifiers in addition to LR. Across the five patch-level datasets, mean ROC-AUC retention relative to full-dimensional embeddings was 98.5% for LR, 98.9% for MLP, and 98.8% for XGBoost. The same encoder pattern was observed across classifiers: histopathology FMs retained 99.2–99.6% of full-

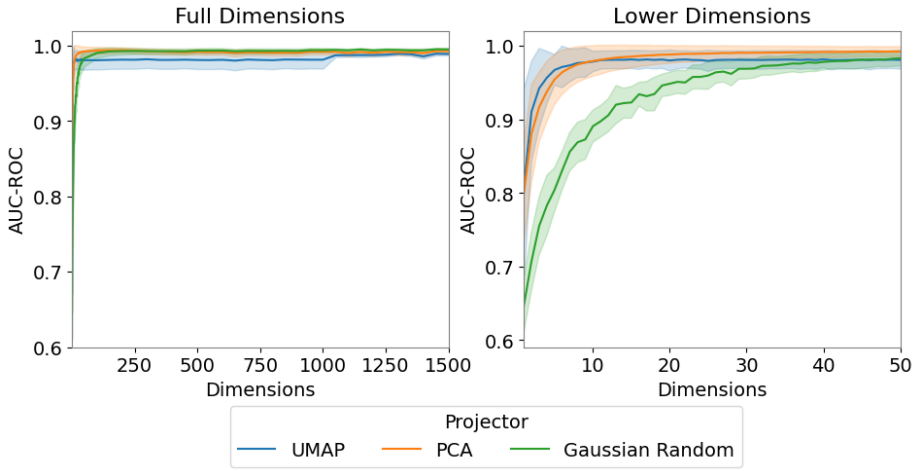

Fig. 1. Aggregate impact of dimensionality reduction on classification performance. The left plot shows aggregate performance from full dimensionality down to 50 dimensions. The right plot zooms into the low-dimensional domain ($k < 50$), where PCA has the highest mean ROC-AUC until approximately 10 dimensions. Solid lines denote the mean ROC-AUC for each projector at each dimension; shaded regions denote ±1 standard deviation across the 20 dataset–encoder pairs and show cross-task dispersion rather than confidence intervals. Individual recovery dimensions are reported in Table 3.

Table 3. Minimum dimensions for 99% ROC-AUC recovery relative to the full-dimensional baseline. d_{ID}: MLE intrinsic dimensionality. P: PCA. U: UMAP. G: GRP. "—": threshold never reached. Full embeddings are 1024-D (UNI, Hibou-L, DINOv2) or 1536-D (UNI2-h). Kather: Kather100K [13]; CRC, Breast, Skin, and Thorax: SPIDER [23].

	UNI				UNI2				Hibou				DINOv2			
Dataset	d_{ID}	P	U	G	d_{ID}	P	U	G	d_{ID}	P	U	G	d_{ID}	P	U	G
Kather	13	5	8	42	13	4	6	33	9	8	—	26	14	11	300	25
CRC	19	5	2	33	20	5	1	33	13	7	2	38	18	23	—	100
Breast	17	10	5	100	17	8	8	100	11	9	25	37	18	100	—	150
Skin	16	11	4	46	17	10	6	40	12	10	8	39	18	36	—	100
Thorax	15	7	4	40	15	8	8	45	11	9	9	37	17	23	—	100

dimensional ROC-AUC on average, while DINOv2 showed larger degradation (95.8–96.7%).

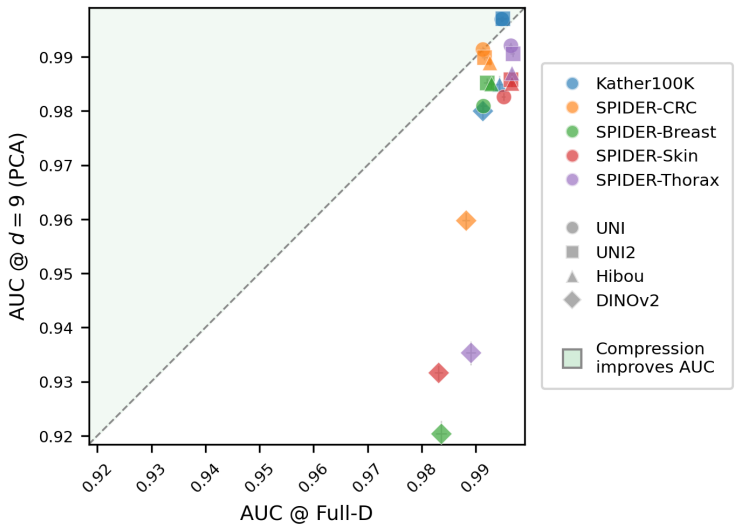

Fig. 2. PCA at $d = 9$. ROC-AUC at PCA $d = 9$ versus full-dimensional ROC-AUC across dataset–encoder pairs. Points near or above the diagonal indicate preserved classification performance.

5 Discussion

Patch-level histopathology FM embeddings contain substantial redundancy: for the median case, PCA retains at least 99% of full-dimensional ROC-AUC in fewer than 10 dimensions, yielding over two orders of magnitude lower dense feature storage for common 1024-D and 1536-D encoders. This is consistent with related work [3]. As a slide-level stress test, we also evaluated the SR386 cohort from SurGen [21], a binary slide-level dataset with 255 training and 84 test slides. The 9-D conclusion did not transfer directly: PCA at $d = 9$ gave LR ROC-AUCs of 0.59–0.66 versus 0.75–0.90 at full dimensionality, while 99% recovery required 20–35 dimensions across encoders. Thus, the strongest claim should remain scoped to patch-level classification; slide-level tasks may still allow large storage reductions, but require task-specific validation.

The linear method, PCA, provides the highest ROC-AUC until single digit dimensions. This may indicate the intrinsic dimensions of the embedding space are sitting on a linear manifold. Since only one nonlinear projector (UMAP) was evaluated, future work should test autoencoders [14] and NCA, and corroborate the manifold hypothesis with additional ID estimators. Histopathology FMs required fewer dimensions to recover 99% ROC-AUC than DINOv2, despite comparable intrinsic dimensionalities; this is consistent with, but does not prove, greater projection robustness from domain-specific pretraining.

PCA at $d = 9$ yields over 100× smaller dense feature arrays for the encoders studied (Table 2), proportionally reducing stored-feature size, CPU–GPU trans-

fer volume, and the input-dependent parameter count of downstream models. These are analytical reductions rather than measured end-to-end speedups, and they do not reduce FM image-encoding time. Lower-dimensional embeddings may also support larger-scale clustering or similarity search.

Limitations. The 99% threshold is descriptive rather than a statistical non-inferiority margin, and representative bootstrap intervals cover selected comparisons rather than every dimension. The study concerns H&E patch classification with ViT-based encoders; CNNs, other modalities, dense prediction, retrieval, and slide-level tasks require separate validation. Reconstruction and neighbourhood preservation were not evaluated; the study measures discriminative information and does not assume invertibility.

6 Conclusion

This work shows that histopathology image embeddings from foundation models can be projected into far fewer dimensions with little loss of performance in the evaluated patch-level classification tasks. PCA provided the strongest low-dimensional performance: 9 dimensions was the median, rather than a universally sufficient target, required to retain 99% of full-dimensional ROC-AUC across the 20 dataset–encoder pairs. This corresponds to over 100× lower dense feature storage for the encoders studied here, subject to task-specific validation.

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

References

1. Ansuini, A., Laio, A., Macke, J.H., Zoccolan, D.: Intrinsic dimension of data representations in deep neural networks (2019)
2. Campanella, G., Chen, S., Singh, M., Verma, R., Muehlstedt, S., Zeng, J., Stock, A., Croken, M., Veremis, B., Elmas, A., Shujski, I., Neittaanmäki, N., Huang, K.l., Kwan, R., Houldsworth, J., Schoenfeld, A.J., Vanderbilt, C.: A clinical benchmark of public self-supervised pathology foundation models. Nature Communications **16**(1) (Apr 2025). https://doi.org/10.1038/s41467-025-58796-1
3. Chandranegara, D.R., Niedziela, P., Cyganek, B.: Compact dino-vit: Feature reduction for visual transformer. Electronics **13**(23) (2024). https://doi.org/10.3390/electronics13234694
4. 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)
5. Chen, T., Guestrin, C.: Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 785–794. KDD '16, ACM (Aug 2016). https://doi.org/10.1145/2939672.2939785
6. Cunningham, J.P., Ghahramani, Z.: Linear dimensionality reduction: Survey, insights, and generalizations. Journal of Machine Learning Research **16**(89), 2859–2900 (2015), https://jmlr.org/papers/v16/cunningham15a.html

7. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: Transformers for image recognition at scale (2021)
8. Filiot, A., Dop, N., Tchita, O., Riou, A., Dubois, R., Peeters, T., Valter, D., Scalbert, M., Saillard, C., Robin, G., Olivier, A.: Distilling foundation models for robust and efficient models in digital pathology. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2025. Lecture Notes in Computer Science, vol. 15966, pp. 162–172. Springer Nature Switzerland (2025). https://doi.org/10.1007/978-3-032-04981-0_16
9. Filiot, A., Jacob, P., Kain, A.M., Saillard, C.: Phikon-v2, a large and public feature extractor for biomarker prediction. ArXiv **abs/2409.09173** (2024)
10. Goldberger, J., Roweis, S., Hinton, G., Salakhutdinov, R.: Neighbourhood components analysis. In: Proceedings of the 18th International Conference on Neural Information Processing Systems. pp. 513–520. NIPS'04, MIT Press, Cambridge, MA, USA (2004)
11. de Jong, E.D., Marcus, E., Teuwen, J.: Current pathology foundation models are unrobust to medical center differences (2025)
12. kaiko.ai, Aben, N., de Jong, E.D., Gatopoulos, I., Kanzig, N., Karasikov, M., Lagr'e, A., Moser, R., van Doorn, J., Tang, F.: Towards large-scale training of pathology foundation models. ArXiv **abs/2404.15217** (2024)
13. Kather, J.N., Halama, N., Marx, A.: 100,000 histological images of human colorectal cancer and healthy tissue (2018). https://doi.org/10.5281/zenodo.1214456
14. LeCun, Y.: Modèles Connexionnistes de l'Apprentissage. Ph.D. thesis, Université Pierre et Marie Curie (Paris 6) (1987)
15. Levina, E., Bickel, P.: Maximum likelihood estimation of intrinsic dimension. Advances in neural information processing systems **17** (2004)
16. Li, H., Shui, Z., Zhang, Y., Zhu, C., Yang, L.F.: Pathvq: Reforming computational pathology foundation model for whole slide image analysis via vector quantization. In: Advances in Neural Information Processing Systems. vol. 38 (2025)
17. Lu, M.Y., Williamson, D.F.K., 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 (Mar 2021). https://doi.org/10.1038/s41551-020-00682-w
18. Mahbod, A., Saeidi, N., Hatamikia, S., Woitek, R.: Evaluating pre-trained convolutional neural networks and foundation models as feature extractors for content-based medical image retrieval. Engineering Applications of Artificial Intelligence **150**, 110571 (Jun 2025). https://doi.org/10.1016/j.engappai.2025.110571
19. Marza, P., Fillioux, L., Boutaj, S., Mahatha, K., Desrosiers, C., Piantanida, P., Dolz, J., Christodoulidis, S., Vakalopoulou, M.: Thunder: Tile-level histopathology image understanding benchmark. In: Advances in Neural Information Processing Systems. vol. 38 (2025), datasets and Benchmarks Track
20. McInnes, L., Healy, J., Melville, J.: Umap: Uniform manifold approximation and projection for dimension reduction (2018), https://arxiv.org/abs/1802.03426
21. Myles, C., Um, I.H., Marshall, C., Harris-Birtill, D., Harrison, D.J.: Surgen: 1020 h&e-stained whole-slide images with survival and genetic markers. GigaScience **14**, giaf086 (2025). https://doi.org/10.1093/gigascience/giaf086
22. Nechaev, D., Pchelnikov, A., Ivanova, E.: Hibou: A family of foundational vision transformers for pathology (2024)
23. Nechaev, D., Pchelnikov, A., Ivanova, E.: Spider: A comprehensive multi-organ supervised pathology dataset and baseline models. arXiv preprint arXiv:2503.02876 (2025)

24. Neidlinger, P., El Nahhas, O.S.M., Muti, H.S., Lenz, T., Hoffmeister, M., Brenner, H., van Treeck, M., Langer, R., Dislich, B., Behrens, H.M., Röcken, C., Foersch, S., Truhn, D., Marra, A., Saldanha, O.L., Kather, J.N.: Benchmarking foundation models as feature extractors for weakly supervised computational pathology. Nature Biomedical Engineering **10**, 1113–1123 (2026). https://doi.org/10.1038/s41551-025-01516-3
25. Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., Assran, M., Ballas, N., Galuba, W., Howes, R., Huang, P.Y., Li, S.W., Misra, I., Rabbat, M., Sharma, V., Synnaeve, G., Xu, H., Jegou, H., Mairal, J., Labatut, P., Joulin, A., Bojanowski, P.: Dinov2: Learning robust visual features without supervision (2024)
26. RAPIDS Development Team: cuML: GPU machine learning algorithms (2024), https://github.com/rapidsai/cuml, version 24.12
27. Vig, P., Wu, R.C., Lotter, W.: Do pathology foundation models encode disease progression? a pseudotime analysis of visual representations. arXiv preprint arXiv:2601.21334 (2026), https://arxiv.org/abs/2601.21334
28. Vorontsov, E., Bozkurt, A., Casson, A., Shaikovski, G., Zelechowski, M., Severson, K., Zimmermann, E., Hall, J., Tenenholtz, N., Fusi, N., Yang, E., Mathieu, P., van Eck, A., Lee, D., Viret, J., Robert, E., Wang, Y.K., Kunz, J.D., Lee, M.C.H., Bernhard, J.H., Godrich, R.A., Oakley, G., Millar, E., Hanna, M.G., Wen, H.Y., Retamero, J., Moye, W.A., Yousfi, R., Kanan, C., Klimstra, D., Rothrock, B., Liu, S., Fuchs, T.J.: A foundation model for clinical-grade computational pathology and rare cancers detection. Nature Medicine **30**, 2924 – 2935 (2024)
29. Wölflein, G., Ferber, D., Meneghetti, A.R., Nahhas, O.S.M.E., Truhn, D., Carrero, Z.I., Harrison, D.J., Arandjelović, O., Kather, J.N.: Benchmarking pathology feature extractors for whole slide image classification (2024)