One-Shot Data Selection for Medical Image Classification via Graph Coverage

Zahiriddin Rustamov[1,2] [0000−0003−4977−1781]*, Nadia Badawi[1] [0009−0004−8795−3521], Rafat Damseh[1] [0000−0001−6797−0448], and Nazar Zaki[1] [0000−0002−6259−9843]

[1] Department of Computer Science and Software Engineering, United Arab Emirates University, Al Ain, United Arab Emirates
{700043167, 700047012, rdamseh, Nzaki}@uaeu.ac.ae
[2] Department of Computer Science, KU Leuven, Leuven, Belgium

Abstract. Training medical image classifiers on entire datasets is wasteful when labeling or compute budgets are constrained: not all samples contribute equally to downstream accuracy. Active learning reduces annotation cost through iterative querying, but assumes repeated access to an oracle and requires multiple rounds of model training. One-shot geometry-based methods such as facility location avoid retraining but operate on pairwise distances that ignore the local structure of the data manifold.

We propose a graph-based one-shot selection method that prunes a labeled training pool using frozen foundation model embeddings. Given embeddings from a pretrained encoder, we construct a k-nearest neighbor graph over all training samples and derive a two-term coverage kernel from the heat diffusion kernel, capturing both direct and two-hop neighborhood relationships. Greedy facility location on this kernel selects class-balanced subsets that maximize coverage of the data manifold. The two-term kernel matches the full spectral heat kernel in selection behavior while reducing computation to sparse matrix operations with a single hyperparameter.

We evaluate on five MedMNIST datasets spanning histopathology, radiology, and microscopy, comparing against both training-dynamics and geometry-based baselines. Our method achieves the highest balanced accuracy on nine of ten dataset-ratio conditions, with the largest gains on class-imbalanced datasets where global graph construction captures cross-class structure that per-class methods miss, all without any model training during selection. Code is available at https://github.com/zahiriddin-rustamov/graph-coverage-selection.

Keywords: Data selection · Coreset · Medical image classification · Graph coverage

* Corresponding author: 700043167@uaeu.ac.ae

1 Introduction

Automated classification of medical images supports clinical workflows from cancer grading in histopathology to disease screening in radiology, but training accurate classifiers requires large annotated datasets. In medical imaging, annotation depends on domain experts whose time is scarce and expensive, and data access is often constrained by institutional policies and patient privacy [1,18]. When labeling or training budgets are limited, the question is not just how many samples to use, but *which* samples to use.

Active learning addresses this through iterative querying: a model is trained, the most informative unlabeled samples are selected, and an expert annotates them [15]. However, it assumes ongoing access to an annotation oracle and requires multiple rounds of training and querying. Annotation budgets are often fixed upfront by study protocols or institutional review, and repeated expert access is impractical; even in cold-start settings, no standard method consistently outperforms random sampling [11]. *One-shot* data selection, where the subset is chosen before any training, avoids these constraints. We target its labeled-pool variant: identifying a small representative subset of an already-labeled training pool that preserves downstream classification accuracy.

Outside medical imaging, one-shot selection has been studied through two families of methods. *Training-dynamics* approaches score samples based on their behavior during training: EL2N [14] uses early-training error norms, forgetting events [17] count classification reversals across epochs, and EVA [8] identifies samples with high cross-epoch variance. These methods are effective [16] but require training a model before any selection occurs, which is circular when the goal is to reduce training cost. *Geometry-based* approaches operate directly on feature representations: facility location [19] selects subsets that maximize coverage in feature space and herding [20] iteratively matches subset statistics to the full dataset. These methods rely on pairwise distances without modeling local manifold structure: two samples may be equidistant from a third but lie in regions of very different density.

The availability of foundation models pretrained on large-scale medical image collections changes what is feasible for one-shot selection. Models such as UNI [2] produce high-quality embeddings without task-specific training. Several methods leverage such embeddings for medical data selection: ENRICH [4] applies diversity sampling on learned features and confounder-aware approaches [9] mitigate confounding variable bias, but both still operate on flat pairwise similarities. Representative Annotation [22] shares the one-shot coverage philosophy for biomedical segmentation but builds per-dataset autoencoder features and a two-stage cluster-then-max-cover, without propagation or class-balance constraints. A k-nearest neighbor graph on these embeddings captures local manifold structure explicitly: propagation through the graph reveals redundancy and coverage relationships that flat distance matrices miss. While graphs are well-established in medical image analysis for *prediction* [23,13,3], modeling *sample-to-sample* relationships via multi-hop graph propagation for classification subset selection is, to our knowledge, under-explored.

We propose a graph-based one-shot selection method for medical image classification. Given frozen UNI embeddings, we build a global k-NN graph over all training samples and compute a coverage kernel $\mathbf{K} = \mathbf{A}_{\text{sym}} + \mathbf{A}_{\text{sym}}^2$ that captures both direct and indirect neighborhood relationships. This kernel arises from decomposition of the heat diffusion kernel; we show that only the first two polynomial terms are necessary, reducing computation to sparse matrix operations. Greedy facility location on this kernel selects diverse, representative subsets with per-class budget constraints.

Our contributions are:

1. A one-shot selection method for medical image classification that operates on frozen foundation model embeddings, requiring no model training before or during selection.
2. Evidence that graph-based selection requires only local adjacency: a two-term polynomial on the normalized adjacency matches the full spectral heat kernel in selection quality, reducing computation from cubic eigendecomposition to linear-time sparse operations.
3. Empirical evidence across five MedMNIST datasets that global graph construction captures decision-boundary information that per-class methods miss, with stronger gains on class-imbalanced datasets.

2 Method

Let $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$ be a labeled training set with C classes, and $f : \mathcal{X} \to \mathbb{R}^d$ a pretrained foundation model encoder. Given a per-class budget B, we select a subset $\mathcal{S} \subset \mathcal{D}$ with $|\mathcal{S}_c| = B$ for each class c to maximize downstream classification accuracy. Selection operates on frozen embeddings $\mathbf{z}_i = f(\mathbf{x}_i)$ (Fig. 1).

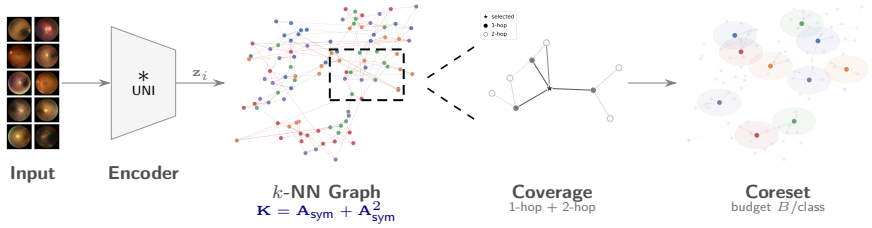

Fig. 1. Overview of graph coverage selection (illustrated on RetinaMNIST). A frozen UNI encoder maps images to embeddings. A k-NN graph is built over all samples, and the coverage kernel $\mathbf{K}$ captures both direct and 2-hop neighborhood structure (center panel: ⋆ selected, ● 1-hop, ○ 2-hop). Greedy facility location selects a balanced coreset maximizing total coverage.

2.1 Neighborhood Graph from Embeddings

The relationship between two samples depends on their shared context: two samples may be equidistant from a third, but one shares a dense local neighborhood with it while the other is isolated. A k-nearest neighbor graph captures this local context by connecting each sample only to its closest neighbors, preserving the manifold structure of the embedding space.

We construct a k-NN graph $\mathcal{G}$ over all N embeddings using cosine similarity. Embeddings are L_2-normalized and each sample is connected to its k nearest neighbors in Euclidean space. Let $\mathbf{A} \in \mathbb{R}^{N \times N}$ denote the weighted adjacency matrix, with edge weights reflecting pairwise similarity. We symmetrize: $\mathbf{A} \leftarrow \max(\mathbf{A}, \mathbf{A}^\top)$, so that neighborhood is a mutual relationship.

2.2 Coverage Kernel via Multi-Hop Propagation

Preliminaries. The heat diffusion kernel $\mathbf{K}_t = \exp(-t\mathbf{L})$ with normalized Laplacian $\mathbf{L} = \mathbf{I} - \mathbf{A}_{\text{sym}}$ propagates similarity along graph edges, with parameter t controlling diffusion locality [10]. Graph coverage measures how well a selected subset's neighborhoods span the sample manifold: when each sample's coverage is its maximum kernel value to any selected sample, greedy maximization of the total prefers diverse, representative subsets.

The adjacency matrix $\mathbf{A}$ encodes direct (1-hop) neighbor relationships. However, two samples that are not direct neighbors may still be redundant if they share many common neighbors. Capturing this requires looking beyond direct connections. We achieve this through a two-step process: symmetric normalization followed by squaring the adjacency. We first normalize $\mathbf{A}$ to account for varying node degrees:

$$\mathbf{A}_{\text{sym}} = \mathbf{D}^{-1/2}\mathbf{A}\,\mathbf{D}^{-1/2}, \tag{1}$$

where $\mathbf{D} = \text{diag}(\mathbf{A1})$ is the degree matrix. This symmetric normalization ensures that if sample i covers sample j, then j equally covers i, which is necessary for consistent coverage accounting during selection. The coverage kernel is then:

$$\mathbf{K} = \mathbf{A}_{\text{sym}} + \mathbf{A}_{\text{sym}}^2. \tag{2}$$

Entry K_{ij} quantifies how well selecting sample j would cover sample i, accounting for both direct proximity and shared neighborhood structure.

Relationship to the heat diffusion kernel. $\mathbf{K}$ is a two-term truncation of the heat-kernel power-series expansion $\mathbf{K}_t = \sum_{k=0}^{\infty} c_k \mathbf{A}_{\text{sym}}^k$ with Poisson coefficients; the k=0 constant offsets every entry equally and does not affect greedy selection. Table 2 confirms that this truncation matches the full spectral kernel in selection behavior.

2.3 Greedy Selection with Per-Class Budget Constraints

Given the kernel $\mathbf{K}$, we select B samples per class to maximize total coverage using greedy facility location [12]. Facility location targets coverage; alternative

submodular families such as graph-cut and log-det target diversity, which is less aligned with low-budget regimes where representative prototypes are preferable to spread. We construct the graph *globally* across all classes rather than independently per class. Per-class graphs can only optimize within-class coverage. A global graph also encodes cross-class proximity: samples near the decision boundary have non-zero kernel entries to samples of other classes, making them more valuable for coverage because selecting them simultaneously represents their own class neighborhood and captures information about the class boundary.

Let $\mathrm{cov}_i = \max_{j \in \mathcal{S}} K_{ij}$ denote the current coverage of sample i by the selected set $\mathcal{S}$. At each step, we select the sample with the largest marginal coverage gain:

$$j^* = \arg \max_{j \notin \mathcal{S},\ |\mathcal{S}_{y_j}| < B} \sum_{i=1}^{N} \max(0,\ K_{ij} - \mathrm{cov}_i), \tag{3}$$

where the constraint $|\mathcal{S}_{y_j}| < B$ enforces that class y_j has not reached its budget. The facility location objective is monotone submodular under per-class budget constraints (a partition matroid), and greedy maximization provides a 1/2-approximation guarantee [5]. The kernel (Eq. 2) and the marginal-gain objective in Eq. 3 are computed entirely from embeddings; class labels enter only through the budget constraint, and removing the constraint leaves the monotone-submodular guarantee intact.

3 Experiments

3.1 Experimental Setup

Datasets. We evaluate on MedMNIST v2 [21], a standardized collection of medical image classification benchmarks. We select five multi-class datasets spanning multiple modalities and scales: **OrganAMNIST** (34,561 training, 11 classes, abdominal CT axial view), **OrganSMNIST** (13,932 training, 11 classes, abdominal CT sagittal view), **PathMNIST** (89,996 training, 9 classes, colorectal cancer histology), **TissueMNIST** (165,466 training, 8 classes, kidney cortex microscopy), and **BloodMNIST** (11,959 training, 8 classes, blood cell microscopy). We use **DermaMNIST** (7,007 training, 7 classes, dermatoscopy) for the global graph analysis due to its severe class imbalance (58:1 ratio).

Embeddings. We use the 224×224 MedMNIST+ variant and extract embeddings using UNI [2], a ViT-L/16 pretrained on over 100 million histopathology images. Images are passed through the frozen encoder, yielding 1024-dimensional embeddings. UNI is a public, large-scale histopathology foundation encoder adopted without task-specific tuning; holding it constant across all selectors isolates the selection mechanism from encoder choice.

Training Protocol. For each selected subset, we train a ResNet-18 [7] from scratch at 224×224 resolution using SGD (learning rate 0.1, momentum 0.9, weight decay 5×10^{-4}) for 1,000 epochs with batch size 256. We use cosine annealing for learning rate scheduling and no data augmentation, to isolate the effect of data selection from training regularization.

Baselines. i) **Random**: balanced random sampling with equal samples per class. *Training-based:* ii) **EL2N** [14]: selects by error L_2-norm from early training epochs. iii) **Forgetting** [17]: selects the most forgotten samples, counting classification reversals across full training. iv) **EVA** [8]: selects samples with high training-error variability across consecutive epochs, requiring 200 epochs of pre-training. *One-shot geometry:* v) **Facility Location** [6]: greedy coverage maximization on full pairwise cosine similarity, the standard geometry-based coreset method. vi) **FPS**: farthest point sampling in embedding space, selecting for maximum spread. vii) **Herding** [20]: iterative mean-matching that selects samples to minimize the distance between subset and full-class centroids in embedding space. All methods use balanced per-class budgets. Each training-dynamics baseline uses its published scoring recipe; the downstream training protocol above is held fixed across all selectors.
Evaluation. We report balanced accuracy. Each configuration is run with 5 random trials (varying training seed), and we report mean ± standard deviation. Selection ratios: 2% and 5% of training data.

3.2 Main Results

Table 1 reports balanced accuracy across five datasets at 2% and 5% selection ratios. Our method achieves the highest balanced accuracy on four of five datasets at 2% and all five at 5%, with the largest gains on PathMNIST (+3.9 percentage points (pp) over facility location at 2%) and BloodMNIST (+5.2 pp at 5%). Among one-shot methods, herding is the closest competitor, trailing by small margins on OrganSMNIST (+0.5 pp) and OrganAMNIST (+0.2 pp at 2%) but falling further behind on PathMNIST and BloodMNIST where local manifold structure is more heterogeneous. Herding operates per-class by design (matching subset centroids to class centroids) and cannot leverage cross-class graph structure (Table 3). Across all ten dataset-ratio conditions, our method outperforms both herding and facility location on nine (sign test, $p = 0.02$); the sole exception is TissueMNIST at 2%, where embedding quality bounds all methods. Facility location, herding, and our method outperform all three training-dynamics baselines on every dataset at both ratios, indicating that at low selection budgets, representation-based coverage is more effective than difficulty-based scoring. On TissueMNIST, where full-data accuracy remains below 70% [21] due to limited discriminability in UNI embeddings for fluorescence microscopy, the coverage-based methods cluster within 2 pp of each other, as selection quality is bounded by representation quality.

3.3 Analysis

Ablation study. Table 2 summarizes ablations on kernel design. No propagation depth wins universally: 1-hop is best on PathMNIST 5%, 2-hop on PathMNIST 2% and OrganAMNIST 5%, and 3-hop on OrganAMNIST 2%, with a maximum spread of 1.6 pp across depths (evaluated at k=5 to amplify hop differences). Coverage in a k-NN graph is inherently local: selecting a sample already "covers"

Table 1. Balanced accuracy (%) on five MedMNIST datasets. **Bold**: best, underline: second best. Mean±std over 5 trials.

			Training-dynamics			*One-shot geometry*			
Dataset	Ratio	Random	EL2N	Forg.	EVA	Facility	FPS	Herding	Ours
OrganS	2%	57.2±4.7	39.5±2.0	42.5±1.4	52.3±1.3	63.3±1.2	59.6±2.7	63.2±0.6	**63.7**±0.7
	5%	66.0±1.2	51.4±0.9	56.6±1.1	66.0±0.9	67.7±1.1	66.0±0.8	68.1±0.7	**68.4**±1.0
OrganA	2%	83.6±2.1	63.1±0.6	72.9±1.8	68.6±2.2	84.9±1.1	83.3±1.5	86.3±0.7	**86.5**±0.9
	5%	89.8±1.2	80.9±0.8	84.6±0.2	81.0±2.0	90.5±0.9	90.7±0.6	90.4±0.7	**91.9**±0.3
Path	2%	77.5±4.6	36.8±1.0	53.5±3.9	53.5±5.7	77.0±1.6	73.0±3.8	78.0±2.1	**80.9**±0.9
	5%	84.0±2.4	49.9±2.4	58.9±4.2	58.7±3.5	82.5±2.2	83.0±1.6	84.6±2.7	**85.9**±1.7
Tissue	2%	43.1±1.0	10.1±0.4	39.3±0.6	39.8±0.6	**44.7**±1.0	35.6±0.9	42.9±0.9	42.6±0.6
	5%	49.1±0.4	15.2±0.2	43.6±1.0	46.7±1.0	48.1±0.6	41.0±1.1	48.3±0.4	**49.5**±1.3
Blood	2%	83.2±1.5	48.3±6.9	67.5±3.3	75.9±2.5	82.7±1.6	78.8±2.1	83.1±3.0	**84.5**±1.7
	5%	90.3±0.5	72.5±1.7	81.4±2.1	86.6±1.7	88.2±2.7	89.0±2.5	92.3±0.6	**93.4**±0.7

its direct neighbors, and propagation beyond two hops reaches samples better served by their own nearby representatives. Replacing our polynomial kernel with the full heat kernel $\exp(-t\mathbf{L})$ yields comparable accuracy and near-identical per-sample coverage scores ($\rho > 0.999$, Table 2), confirming that three components of the heat-kernel expansion are unnecessary for selection: eigendecomposition (the polynomial form suffices), terms beyond $\mathbf{A}^2_{\text{sym}}$ (negligible contribution), and Poisson decay coefficients (no benefit over uniform weighting). This reduces computation from $O(N^3)$ eigendecomposition to $O(Nk^2)$ sparse matrix multiplication, leaving k as the sole hyperparameter. Performance generally improves with graph density and stabilizes around k=20, though BloodMNIST at 2% continues to benefit from denser graphs (k=50). Very sparse graphs (k=5) underperform on BloodMNIST and OrganSMNIST at 2%, suggesting that low budgets require sufficient neighborhood structure. The sparse kernel is also necessary at scale: on TissueMNIST (165K samples), facility location on the full pairwise matrix exceeds GPU memory (102 GB), while graph construction and selection complete in 89 seconds.

Global vs. per-class graph construction. Table 3 and Fig. 2 compare global and per-class graph construction on DermaMNIST, which exhibits severe class imbalance (58:1 ratio). Global construction improves balanced accuracy by 5.7 pp at 2% budget. The mechanism is visible in Fig. 2: per-class graphs for minority classes contain so few nodes that they are nearly complete, leaving little room for selection to improve coverage beyond random. Global construction connects minority-class samples to the broader manifold, so the greedy algorithm selects representatives near decision boundaries rather than redundant intra-class neighbors. Herding, which is inherently per-class, achieves 39.6% on DermaMNIST at 2%, closer to our per-class baseline (38.1%) than to global construction (43.8%), suggesting that methods without cross-class structure cannot exploit this advantage. The gap narrows at 5% (+3.4 pp), consistent with minority classes having enough budget to cover their own regions at higher ratios. The same pattern

Table 2. Ablation study. Balanced accuracy (%) under different kernel configurations.

	PathMNIST		OrganAMNIST	
	2%	5%	2%	5%
Propagation depth				
$\mathbf{A}$ (1-hop)	$79.4_{\pm 2.2}$	$85.4_{\pm 2.3}$	$86.7_{\pm 0.6}$	$91.0_{\pm 0.6}$
$\mathbf{A}+\mathbf{A}^2$ (2-hop)	$81.0_{\pm 2.3}$	$83.9_{\pm 1.4}$	$87.1_{\pm 0.2}$	$91.2_{\pm 0.4}$
$\mathbf{A}+\mathbf{A}^2+\mathbf{A}^3$ (3-hop)	$80.3_{\pm 0.8}$	$84.5_{\pm 2.1}$	$87.7_{\pm 0.3}$	$90.6_{\pm 1.1}$
	BloodMNIST		OrganSMNIST	
	2%	5%	2%	5%
Kernel type				
Polynomial ($\mathbf{A}+\mathbf{A}^2$)	$84.5_{\pm 1.7}$	$93.4_{\pm 0.7}$	$63.6_{\pm 0.7}$	$68.4_{\pm 1.0}$
Heat kernel ($e^{-t\mathbf{L}}$)	$84.6_{\pm 1.5}$	$93.2_{\pm 0.7}$	$62.4_{\pm 0.6}$	$68.0_{\pm 0.5}$
Coverage ρ	0.999		0.999	
Neighborhood size				
$k = 5$	$81.7_{\pm 2.4}$	$92.5_{\pm 1.1}$	$60.1_{\pm 1.1}$	$68.1_{\pm 0.5}$
$k = 20$	$81.9_{\pm 1.8}$	$92.5_{\pm 1.2}$	$63.6_{\pm 0.7}$	$68.0_{\pm 0.5}$
$k = 50$	$84.5_{\pm 1.7}$	$93.4_{\pm 0.7}$	$63.7_{\pm 0.7}$	$68.4_{\pm 1.0}$

Table 3. Global vs. per-class graph on DermaMNIST.

Scope	2%	5%
Per-class	$38.1_{\pm 1.3}$	$48.9_{\pm 1.6}$
Global	$\mathbf{43.8}_{\pm 2.6}$	$\mathbf{52.3}_{\pm 1.9}$
Δ	+5.7	+3.4

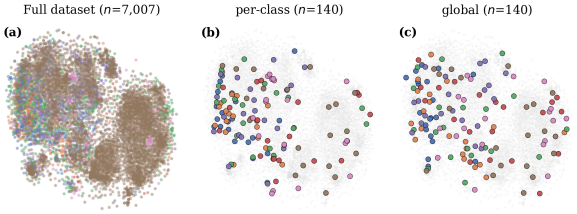

Fig. 2. Effect of global graph construction on DermaMNIST. (a) Full dataset in t-SNE space; the majority class dominates. (b) Per-class selection clusters within individual class regions. (c) Global selection distributes samples into cross-class overlap zones.

holds on BloodMNIST (84.5% vs. 83.6% at 2%) and OrganSMNIST (63.7% vs. 62.5%).

Training-dynamics methods. EL2N, forgetting, and EVA consistently underperform random selection on nearly every dataset and ratio (Table 1). EL2N achieves 36.8% on PathMNIST at 2% compared to 77.5% for random, and 10.1% on TissueMNIST (near chance for 8 classes). These methods select hard or ambiguous samples, which are informative for pruning large datasets but counterproductive at low budgets where the model needs representative prototypes, not boundary cases [16].

4 Conclusion

We presented a one-shot data selection method for medical image classification that builds a k-NN graph on frozen foundation model embeddings and selects subsets via greedy facility location on a two-term coverage kernel. The approach requires no model training before or during selection, and a single hyperparameter.

Three findings emerge from our experiments. First, the value lies in graph structure, not propagation sophistication: the k-NN adjacency and its square capture nearly all selection-relevant information from the full spectral heat kernel, making eigendecomposition and decay coefficients unnecessary. Second, global graph construction across classes consistently outperforms per-class construction, with the largest gains under class imbalance where minority-class graphs are too small for meaningful selection. Third, training-dynamics methods are counterproductive at low budgets: they were designed for pruning large datasets, not selecting the 2–5% of data where representative prototypes matter most.

The main limitation is dependence on embedding quality. On TissueMNIST, where UNI embeddings poorly separate fluorescence microscopy classes, all selection methods converge, as graph structure cannot compensate for uninformative representations. Extending this approach to other tasks, evaluating with domain-specific foundation models beyond pathology, scaling to high-resolution clinical images and cross-institution domain shift, conducting direct boundary-coverage analyses, and developing a fully label-free variant (e.g., with per-class budgets estimated from cluster structure) remain open directions.

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

References

1. Budd, S., Robinson, E.C., Kainz, B.: A survey on active learning and human-in-the-loop deep learning for medical image analysis. Medical Image Analysis **71**, 102062 (2021)
2. Chen, R.J., Ding, T., Lu, M.Y., Williamson, D.F.K., Jaume, G., Song, A.H., Chen, B., Zhang, A., Shao, D., Shaban, M., Williams, M., Oldenburg, L., Weishaupt, L.L., Wang, J.J., Vaidya, A., Le, L.P., Gerber, G., Sahai, S., Williams, W., Mahmood, F.: Towards a general-purpose foundation model for computational pathology. Nature Medicine **30**(3), 850–862 (2024)
3. Chen, R.J., Lu, M.Y., Shaban, M., Chen, C., Chen, T.Y., Williamson, D.F.K., Mahmood, F.: Whole slide images are 2d point clouds: Context-aware survival prediction using patch-based graph convolutional networks. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. pp. 339–349. Springer (2021)
4. Chinn, E., Arora, R., Arnaout, R., Arnaout, R.: Enriching medical imaging training sets enables more efficient machine learning. Journal of the American Medical Informatics Association **30**(6), 1079–1090 (2023). https://doi.org/10.1093/jamia/ocad055

5. Fisher, M.L., Nemhauser, G.L., Wolsey, L.A.: An analysis of approximations for maximizing submodular set functions—II. In: Balinski, M.L., Hoffman, A.J. (eds.) Polyhedral Combinatorics, pp. 73–87. Springer, Berlin, Heidelberg (1978)
6. Guo, C., Zhao, B., Bai, Y.: Deepcore: A comprehensive library for coreset selection in deep learning. In: DEXA. LNCS, vol. 13108, pp. 181–195. Springer (2022)
7. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR (2016)
8. Hong, Y., Zhang, X., Zhang, X., Zhou, J.T.: Evolution-aware variance (eva) coreset selection for medical image classification. In: Proceedings of the 32nd ACM International Conference on Multimedia. pp. 301–310. MM '24, Association for Computing Machinery, New York, NY, USA (2024). https://doi.org/10.1145/3664647.3681592
9. Ji, A., Kang, Q., Xu, W., Wang, C., Li, K., Lao, Q.: Confounder-aware medical data selection for fine-tuning pretrained vision models. In: 2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI) (2025). https://doi.org/10.1109/ISBI60581.2025.10980785
10. Kondor, R.I., Lafferty, J.: Diffusion kernels on graphs and other discrete input spaces. In: ICML (2002)
11. Liu, H., Li, H., Yao, X., Fan, Y., Hu, D., Dawant, B.M., Nath, V., Xu, Z., Oguz, I.: Colossal: A benchmark for cold-start active learning for 3d medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. pp. 25–34. Springer (2023)
12. Nemhauser, G.L., Wolsey, L.A., Fisher, M.L.: An analysis of approximations for maximizing submodular set functions—I. Mathematical Programming **14**, 265–294 (1978)
13. Pati, P., Jaume, G., Foncubierta-Rodríguez, A., Feroce, F., Anniciello, A.M., Scognamiglio, G., Brancati, N., Fiche, M., Dubruc, E., Riccio, D., Di Bonito, M., De Pietro, G., Botti, G., Thiran, J.P., Frucci, M., Goksel, O., Gabrani, M.: Hierarchical graph representations in digital pathology. Medical Image Analysis **75**, 102264 (2022)
14. Paul, M., Ganguli, S., Dziugaite, G.K.: Deep learning on a data diet: Finding important examples early in training. In: Advances in Neural Information Processing Systems. vol. 34, pp. 20596–20607 (2021)
15. Settles, B.: Active learning literature survey. Tech. rep., University of Wisconsin–Madison (2009)
16. Sorscher, B., Geirhos, R., Shekhar, S., Ganguli, S., Morcos, A.: Beyond neural scaling laws: Beating power law scaling via data pruning. In: Advances in Neural Information Processing Systems. vol. 35, pp. 19523–19536. Curran Associates, Inc. (2022)
17. Toneva, M., Sordoni, A., Combes, R.T.d., Trischler, A., Bengio, Y., Gordon, G.J.: An empirical study of example forgetting during deep neural network learning. In: ICLR (2019)
18. Wang, H., Jin, Q., Li, S., Liu, S., Wang, M., Song, Z.: A comprehensive survey on deep active learning in medical image analysis. Medical Image Analysis **95**, 103201 (2024)
19. Wei, K., Iyer, R., Bilmes, J.: Submodularity in data subset selection and active learning. In: ICML. vol. 37, pp. 1954–1963. PMLR (2015)
20. Welling, M.: Herding dynamical weights to learn. In: ICML (2009)
21. Yang, J., Shi, R., Wei, D., Liu, Z., Zhao, L., Ke, B., Pfister, H., Ni, B.: Medmnist v2–a large-scale lightweight benchmark for 2d and 3d biomedical image classification. Scientific Data **10**(1), 41 (2023)

22. Zheng, H., Yang, L., Chen, J., Han, J., Zhang, Y., Liang, P., Zhao, Z., Wang, C., Chen, D.Z.: Biomedical image segmentation via representative annotation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 33, pp. 5901–5908 (2019). https://doi.org/10.1609/aaai.v33i01.33015901
23. Zhou, Y., Graham, S., Alemi Koohbanani, N., Shaban, M., Heng, P.A., Rajpoot, N.: Cgc-net: Cell graph convolutional network for grading of colorectal cancer histology images. In: ICCV Workshops (2019)