Bio-Inspired Multi-Objective Batch Active Learning for Skin Lesion Classification

Nancy Kalaj
University of Trento
Trento, Italy
nancy.kalaj@studenti.unitn.it

Matteo Bando
University of Trento
Trento, Italy
matteo.bando@studenti.unitn.it

Šimon Rokůsek
University of Trento
Trento, Italy
simon.rokusek@studenti.unitn.it

Ludovico Cappellato
University of Trento
Trento, Italy
ludovico.cappellato@studenti.unitn.it

Erik Nielsen
University of Trento
Trento, Italy
erik.nielsen@unitn.it

Stefano Genetti
University of Trento
Trento, Italy
stefano.genetti@unitn.it

Giovanni Iacca
University of Trento
Trento, Italy
giovanni.iacca@unitn.it

Abstract

Deep learning models for medical image classification achieve high accuracy but rely on large, costly labeled datasets. Active Learning (AL) mitigates this by iteratively selecting informative samples under a fixed labeling budget. However, batch-mode AL often struggles with redundancy when relying solely on scalar uncertainty scores. In this work, we formulate batch-mode AL for skin lesion classification as a Multi-Objective Optimization (MOO) problem, explicitly balancing informativeness, coverage, and diversity. We employ bio-inspired optimizers, NSGA-II with set-based encoding and MOPSO with continuous scoring vectors, to search for Pareto-optimal batches. We evaluated our MOO approaches on the HAM10000 dataset, where they were able to outperform established baselines in sample efficiency, with MOPSO attaining the highest Area Under the Learning Curve (AULC). Furthermore, Pareto-front analyses provide interpretable insights into the trade-offs between the competing objectives. This work demonstrates that evolutionary multi-objective algorithms offer a superior, transparent approach to resolving conflicting acquisition criteria in deep Active Learning.

CCS Concepts

• **Theory of computation** → **Active learning**; *Bio-inspired optimization*; • **Computing methodologies** → **Active learning settings**.

Keywords

Active Learning, Multi-objective Optimization, NSGA-II, MOPSO, Medical Imaging, Deep Learning

GECCO Companion '26, San Jose, Costa Rica

ACM ISBN 979-8-4007-2488-6/2026/07
https://doi.org/10.1145/3795101.3814666

ACM Reference Format:
Nancy Kalaj, Matteo Bando, Šimon Rokůsek, Ludovico Cappellato, Erik Nielsen, Stefano Genetti, and Giovanni Iacca. 2026. Bio-Inspired Multi-Objective Batch Active Learning for Skin Lesion Classification. In *Genetic and Evolutionary Computation Conference (GECCO Companion '26), July 13–17, 2026, San Jose, Costa Rica.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3795101.3814666

1 Introduction

Deep learning has enabled substantial progress in medical image classification, including dermatology. However, these models typically require large labeled datasets, and obtaining annotations for medical images is costly, time-consuming, and dependent on domain experts. This labeling bottleneck is particularly critical in settings such as dermoscopic analysis, where the variability from observer to observer can be significant.

Active Learning (AL) offers a practical solution to this issue by iteratively selecting the most informative unlabeled samples for annotation under a fixed labeling budget. In pool-based AL, a model is trained on a small labeled set and then used to query additional samples from a large unlabeled pool. While classical uncertainty-based strategies (e.g., entropy sampling) are simple and effective, they often perform poorly in batch-mode settings, where multiple samples must be selected simultaneously. In such cases, uncertainty alone can lead to redundant selections, as similar samples tend to share similar uncertainty scores. To address this limitation, modern batch-mode AL methods incorporate additional criteria such as diversity and representativeness, encouraging the selected batch to better cover the underlying data distribution. Existing approaches typically combine these criteria heuristically, e.g., via two-stage pipelines or hand-tuned scalarization of multiple signals. While effective, these designs often do not show the trade-offs between competing objectives and require careful tuning.

In this work, we explicitly formulate batch selection in AL as a multi-objective optimization (MOO) problem. We leverage NSGA-II and MOPSO to search for Pareto-optimal batches, providing a principled mechanism for handling conflicting objectives (Figure 1).

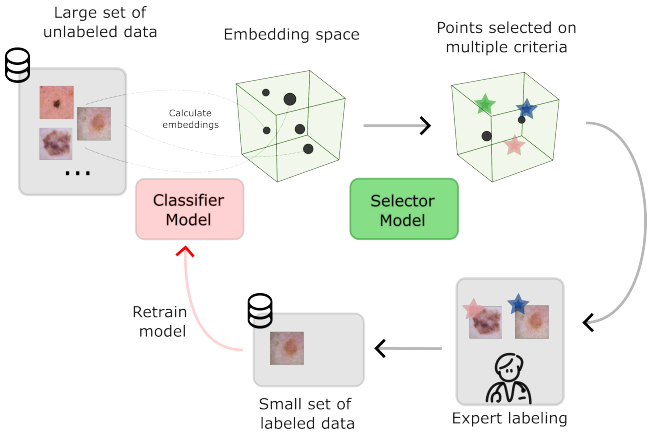


Figure 1: Overview of the proposed approach.

Evaluated on the HAM10000 dataset, our results show that these bio-inspired approaches improve sample efficiency across the learning trajectory while offering an interpretable view of the acquisition process. We make our code publicly available on GitHub[1].

2 Related Work

AL aims to reduce annotation costs by selecting informative samples. Query strategies typically fall into uncertainty-based, representativeness-based, and hybrid approaches [4]. Uncertainty sampling (e.g., predictive entropy [11]) selects low-confidence samples but often yields redundant batches. To mitigate this, representativeness criteria (e.g., Core-Set [7], density-weighting [2]) encourage geometric or distributional coverage but may underexploit uncertainty. Modern hybrid methods combine uncertainty with diversity or representativeness, often via staged pipelines or heuristic scoring. Examples include Suggestive Annotation [12], MedAL [9], BADGE [1], and BatchBALD [3]. While effective, these approaches typically rely on ad hoc combinations of multiple criteria. Alternatively, recent work formulates AL as a multi-objective optimization problem, explicitly balancing competing criteria such as informativeness, coverage, and diversity [5, 6, 13]. These methods offer a principled and interpretable framework but remain relatively underexplored in deep AL, particularly in medical imaging. In contrast to prior work relying on heuristic combinations, we adopt a multi-objective formulation and investigate bio-inspired optimizers to explicitly model trade-offs in batch selection.

3 Problem Formulation

We consider pool-based AL for multi-class image classification. Let $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N}$ be a dataset with inputs $x_i \in \mathcal{X}$ and labels $y_i \in \{1, \dots, C\}$. At round t, the data are split into a labeled set $\mathcal{L}_t$ and an unlabeled pool $\mathcal{U}_t$, with $\mathcal{L}_t \cap \mathcal{U}_t = \emptyset$. A model f_{θ_t} is trained on $\mathcal{L}_t$, and an acquisition function selects a batch $\mathcal{B}_t \subset \mathcal{U}_t$ of size $|\mathcal{B}_t| = B$ for annotation. Let y be the newly acquired vector of labels of size B for the samples in $\mathcal{B}_t$. The sets are updated as:

$$\mathcal{L}_{t+1} = \mathcal{L}_t \cup [\mathcal{B}_t | y], \quad \mathcal{U}_{t+1} = \mathcal{U}_t \setminus \mathcal{B}_t. \tag{1}$$

[1]https://github.com/DIOL-UniTN/mo-batch-active-learning.git

This process is repeated for a fixed number of rounds or until a labeling budget is exhausted. The goal is to maximize model performance under a limited annotation budget.

4 Multi-Objective Batch Selection

We design acquisition strategies that explicitly balance three criteria: informativeness, coverage, and diversity.

Pointwise signals. For each $x \in \mathcal{U}_t$, we compute: (i) predictive entropy $u_t(x)$ as an uncertainty measure, (ii) distance to the labeled set $\delta_t(x)$ in embedding space as a coverage/novelty proxy, and (iii) a density-based representativeness score $r_t(x)$ derived from nearest-neighbor distances within $\mathcal{U}_t$. All signals are normalized within the current pool $\mathcal{U}_t$; we denote normalized quantities with a tilde (e.g., $\tilde{u}_t(x)$, $\tilde{\delta}_t(x)$, $\tilde{r}_t(x)$). Note that, while uncertainty and coverage are used in our batch-level multi-objective formulation, $r_t(x)$ is computed specifically for our hybrid baseline (Section 5).

Multi-objective batch selection. To directly optimize set-level properties, we formulate batch selection as a multi-objective problem over subsets $\mathcal{B} \subseteq \mathcal{U}_t$, $|\mathcal{B}| = B$. Let $z_{\theta_t}(x)$ denote the embedding of sample x, given by the penultimate layer of the network at round t. For a candidate batch $\mathcal{B}$, we define:

$$F_1(\mathcal{B}) = \frac{1}{B} \sum_{x \in \mathcal{B}} \tilde{u}_t(x) \quad \text{(informativeness)}, \tag{2}$$

$$F_2(\mathcal{B}) = \frac{1}{B} \sum_{x \in \mathcal{B}} \tilde{\delta}_t(x) \quad \text{(coverage)}, \tag{3}$$

$$F_3(\mathcal{B}) = \frac{2}{B(B-1)} \sum_{i<j} ||z_{\theta_t}(x_i) - z_{\theta_t}(x_j)||_2 \quad \text{(diversity)}. \tag{4}$$

These objectives are generally conflicting; we therefore seek Pareto-optimal batches rather than a single scalar optimum.

Candidate shortlisting. For efficiency, optimization is performed on a reduced candidate pool $\mathcal{U}_t^{(M)} \subset \mathcal{U}_t$ formed by selecting the top-M samples according to uncertainty. This step reduces the search space while retaining informative candidates.

NSGA-II Representation. We apply NSGA-II with a set-based encoding, where each individual represents a batch of B samples. Individuals are evaluated using (F_1, F_2, F_3) and ranked via non-dominated sorting, with crowding distance promoting diversity along the Pareto front. Variation operators include set crossover and mutation with uniqueness constraints. After a fixed number of generations, a single batch is selected from the Pareto set using a normalized objective sum.

MOPSO Representation. We also employ a Multi-Objective Particle Swarm Optimizer (MOPSO), where each particle encodes a continuous scoring vector over candidate samples. A batch is obtained by selecting the top-B entries. Particles evolve using personal bests and a Pareto archive of non-dominated solutions, with a linearly decreasing inertia weight [8]. The final batch is selected using a knee-point criterion to balance objectives.

5 Experimental Setup

Dataset & Model. We evaluate on the HAM10000 dataset [10], which consists of dermoscopic images from multiple diagnostic

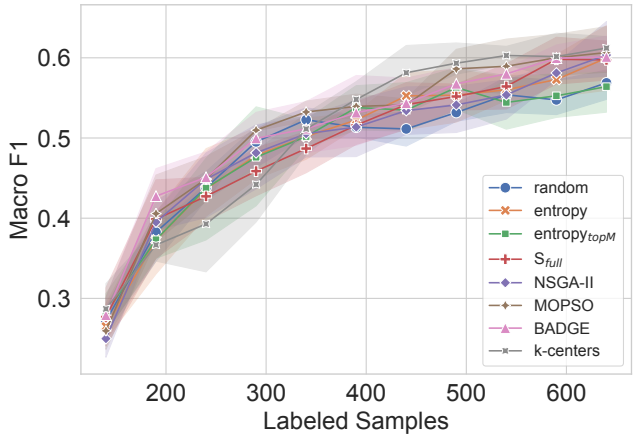


Figure 2: Validation macro-F1 versus labeling budget on HAM10000 (mean ± std, $n=4$ seeds).

categories. We use an ImageNet-pretrained CNN backbone with a classification head adapted for the target classes. Training follows a *freeze-unfreeze* protocol: the backbone is initially frozen while training the head, and then the full network is fine-tuned. At each AL round, the model is retrained from the current labeled set.

Baselines. We compare our approach against random sampling (Random), BADGE [1], entropy-based uncertainty sampling (Entropy), and k-centers (Core-Set) selection [7]. We also include a shortlist baseline (entropy$_{\text{topM}}$) that selects from the same candidate pool used by our multi-objective methods. As a strong hybrid baseline, we include a greedy scoring function (S$_{\text{full}}$) that combines normalized uncertainty, distance-to-labeled (coverage), and representativeness as $S_t(x) = \alpha\tilde{u}_t(x) + \beta\tilde{\delta}_t(x) + \gamma\tilde{r}_t(x)$, and selects the top-$B$ samples. This method incorporates the same signals as our multi-objective formulation but performs independent (pointwise) selection without explicitly enforcing diversity at the batch level.

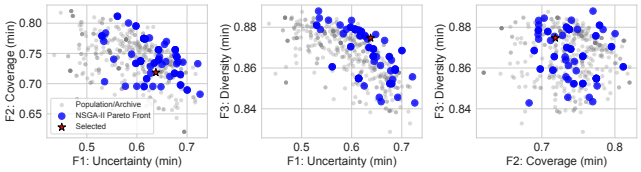
Active learning protocol. We perform pool-based AL for a fixed number of rounds. At each round, a batch of size B is selected from the unlabeled pool and added to the labeled set. All methods use the same initialization, batch size, and training procedure to ensure fair comparison. Results are averaged over four random seeds.

Table 1: Overall performance (mean ± std, $n=4$ seeds).

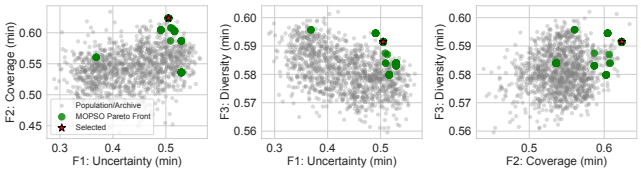
Method	Final macro-F1 @ 640	AULC
Random	0.5687 ± 0.0227	245.85 ± 8.09
Entropy	0.5997 ± 0.0468	249.17 ± 11.09
entropy$_{\text{topM}}$	0.5641 ± 0.0368	247.20 ± 12.30
S$_{\text{full}}$	0.5974 ± 0.0281	249.03 ± 12.33
NSGA-II	0.6035 ± 0.0395	249.05 ± 11.00
MOPSO	0.6065 ± 0.0313	**259.32 ± 6.82**
BADGE	0.6010 ± 0.0270	257.58 ± 11.85
k-centers	**0.6121 ± 0.0149**	254.47 ± 13.28

Table 2: Condensed batch diagnostics (mean ± std, $n=4$ seeds).

Method	Mean entropy	Mean dist-to-labeled	Mean pairwise dist
Random	0.97 ± 0.24	0.595 ± 0.025	0.383 ± 0.016
Entropy	1.76 ± 0.08	0.694 ± 0.022	0.371 ± 0.016
entropy$_{\text{topM}}$	1.76 ± 0.09	0.691 ± 0.021	0.374 ± 0.020
S$_{\text{full}}$	1.63 ± 0.10	0.723 ± 0.024	0.394 ± 0.013
NSGA-II	1.70 ± 0.10	0.710 ± 0.027	0.393 ± 0.013
MOPSO	1.65 ± 0.11	0.701 ± 0.021	0.399 ± 0.012
BADGE	1.39 ± 0.12	0.654 ± 0.021	0.403 ± 0.012
k-centers	1.49 ± 0.12	0.751 ± 0.030	0.418 ± 0.023



(a) NSGA-II



(b) MOPSO

Figure 3: Explored solutions (grey), Pareto front (blue/green), and selected batch (red star) during the initial AL round.

6 Results and Discussion

Figure 2 shows validation macro-F1 versus labeling budget, and Table 1 summarizes results across seeds. At the maximum budget, k-centers achieves the best final macro-F1, while MOPSO and NSGA-II perform comparably to strong baselines. Of note, MOPSO attains the highest AULC (Table 1), indicating superior sample efficiency across budgets. This suggests that multi-objective optimization is particularly beneficial in early and intermediate stages of AL.

Effect of multi-objective selection. Restricting selection to a shortlist of uncertain samples degrades performance (see entropy$_{\text{topM}}$ vs. entropy in Table 1). However, NSGA-II and MOPSO recover this gap by optimizing coverage and diversity within the shortlist. As shown in Figure 2, both methods consistently outperform entropy$_{\text{topM}}$, supporting the view of multi-objective optimization as an effective *batch refinement* step.

Acquisition behavior. Batch-level statistics (Table 2) show that entropy-based methods yield the highest uncertainty but lower diversity, while k-centers maximizes coverage and spread. Multi-objective methods balance all three criteria, explaining their strong AULC performance.

Pareto front analysis. Pareto-front visualizations (Figure 3) show that NSGA-II explores a broader range of trade-offs, while MOPSO concentrates on balanced high-performing regions. In both cases, selected solutions lie near the knee of the Pareto front, achieving a balanced compromise among objectives.

Computational cost. NSGA-II and MOPSO introduce moderate overhead (~1.3× Entropy), while S_{full} is the most expensive due to repeated embedding computations. Overall, MOPSO provides the best trade-off between cost and sample efficiency.

Summary. Overall, multi-objective optimization, particularly MOPSO, improves sample efficiency while maintaining competitive final performance and provides an interpretable framework for balancing uncertainty, coverage, and diversity.

7 Ablation

We analyze three key components: (i) hybrid scoring, (ii) candidate pool size M, and (iii) Pareto decision rules.

(i) Greedy hybrid scoring. Decomposing S_{full} into uncertainty-only and partial variants shows that incorporating embedding-space geometry improves sample efficiency beyond entropy sampling. This confirms the effectiveness of the S_{full} baseline. Detailed results are available in the public codebase.

(ii) Candidate pool size. We study the effect of the shortlist size M (Figure 4, Table 3). Performance is sensitive to M, reflecting a trade-off between candidate restriction and search flexibility. Small pools (e.g., $M = 50$) yield the greatest improvements over the matched `entropy_topM` baseline (in both AULC and final F1), while larger values ($M \geq 500$) provide no consistent benefit and increase computational cost.

(iii) Pareto decision rules. We compare strategies for selecting a solution from the Pareto front (Table 4). The choice of rule materially affects performance: normalized-sum (`sum`) is effective for smaller M, while distance-to-ideal (`ideal`) performs better for larger M. Complex strategies do not yield consistent improvements.

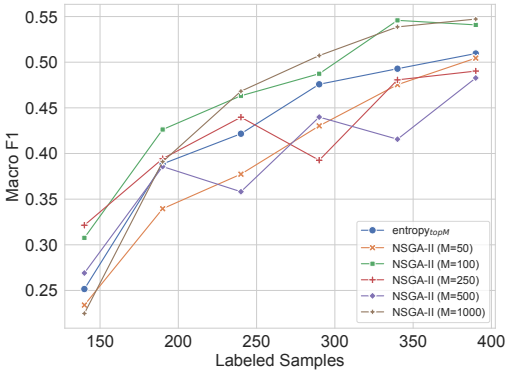


Figure 4: Ablation studies investigating the sensitivity to population size M.

8 Conclusion

We studied pool-based AL for medical image classification and proposed to formulate batch selection as a multi-objective optimization problem balancing uncertainty, coverage, and diversity. Using bio-inspired optimizers (NSGA-II and MOPSO), we showed that it is possible to obtain competitive performance with strong baselines while improving sample efficiency across labeling budgets. In particular, MOPSO achieves the highest AULC, indicating more consistent gains throughout the AL process, while k-centers remains strong at large budgets. Beyond performance, the proposed formulation provides an interpretable view of acquisition behavior through Pareto fronts, making trade-offs between objectives explicit. Overall, our results suggest that multi-objective optimization is a practical and flexible framework for batch-mode AL. Future work includes scaling to larger candidate pools, improving efficiency, and extending the approach to other medical imaging tasks and model architectures.

Table 3: NSGA-II vs. entropy$_{\text{topM}}$ across candidate pool sizes M. Positive values indicate improvements of NSGA-II wrt entropy$_{\text{topM}}$.

M	ΔAULC	ΔF1
50	**+8.76**	**+0.049**
100	+1.77	-0.012
250	-16.98	-0.065
500	-0.24	+0.031
1000	-1.06	-0.012

Table 4: AULC for NSGA-II Pareto decision rules (higher is better).

Chooser	AULC ($M = 100$)	AULC ($M = 250$)
`sum`	118.95	103.44
`ideal`	114.85	117.33
`maximin`	105.07	114.87
`adaptive`	104.00	99.41

References

[1] Ash, Jordan T and Zhang, Chicheng and Krishnamurthy, Akshay and Langford, John and Agarwal, Alekh. 2019. Deep batch active learning by diverse, uncertain gradient lower bounds. https://doi.org/10.48550/arXiv.1906.03671

[2] Budd, Samuel and Robinson, Emma C and Kainz, Bernhard. 2021. A survey on active learning and human-in-the-loop deep learning for medical image analysis. *Medical image analysis* 71 (2021), 102062. https://doi.org/10.1016/j.media.2021.102062

[3] Andreas Kirsch, Joost van Amersfoort, and Yarin Gal. 2019. BatchBALD: Efficient and Diverse Batch Acquisition for Deep Bayesian Active Learning. https://proceedings.neurips.cc/paper_files/paper/2019/file/95323660ed2124450caaac2c46b5ed90-Paper.pdf

[4] Li, Dongyuan and Wang, Zhen and Chen, Yankai and Jiang, Renhe and Ding, Weiping and Okumura, Manabu. 2024. A survey on deep active learning: Recent advances and new frontiers. *IEEE Transactions on Neural Networks and Learning Systems* 36, 4 (2024), 5879–5899. https://doi.org/10.1109/TNNLS.2024.3396463

[5] Mandal, Sayantani and Ghosh, Subhayu and Jana, Nanda Dulal and Chakraborty, Somenath and Mallik, Saurav. 2025. Active learning with particle swarm optimization for enhanced skin cancer classification utilizing deep CNN models. *Journal of Imaging Informatics in Medicine* 38, 4 (2025), 2472–2489. https://doi.org/10.1007/s10278-024-01327-z

[6] Reyes, Oscar and Ventura, Sebastián. 2018. Evolutionary strategy to perform batch-mode active learning on multi-label data. *ACM Transactions on Intelligent Systems and Technology (TIST)* 9, 4 (2018), 1–26. https://doi.org/10.1145/3161606

[7] Sener, Ozan and Savarese, Silvio. 2017. Active learning for convolutional neural networks: A core-set approach. https://doi.org/10.48550/arXiv.1708.00489

[8] Shi, Yuhui and Eberhart, Russell C. 1999. Empirical study of particle swarm optimization. 1945–1950 pages. https://doi.org/10.1109/CEC.1999.785511

[9] Smailagic, Asim and Noh, Hae Young and Costa, Pedro and Walawalkar, Devesh and Khandelwal, Kartik and Mirshekari, Mostafa and Fagert, Jonathon and Galdrán, Adrián and Xu, Susu. 2018. Medal: Deep active learning sampling method for medical image analysis. https://doi.org/10.48550/arXiv.1809.09287

[10] Tschandl, Philipp and Rosendahl, Cliff and Kittler, Harald. 2018. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific data* 5, 1 (2018), 180161. https://doi.org/10.1038/sdata.2018.161

[11] Xie, Binhui and Yuan, Longhui and Li, Shuang and Liu, Chi Harold and Cheng, Xinjing and Wang, Guoren. 2022. Active learning for domain adaptation: An energy-based approach. 8708–8716 pages. https://doi.org/10.1609/aaai.v36i8.20850

[12] Yang, Lin and Zhang, Yizhe and Chen, Jianxu and Zhang, Siyuan and Chen, Danny Z. 2017. Suggestive annotation: A deep active learning framework for biomedical image segmentation. 399–407 pages. https://doi.org/10.1007/978-3-319-66179-7_46

[13] Zhan, Xueying and Dai, Zeyu and Wang, Qingzhong and Li, Qing and Xiong, Haoyi and Dou, Dejing and Chan, Antoni B. 2022. Pareto optimization for active learning under out-of-distribution data scenarios. https://doi.org/10.48550/arXiv.2207.01190