Hard-First: Entropy-Guided Curriculum Distillation Balances Transfer and Preservation in Biomedical Vision-Language Models

Junseob Kim [1] **Sunil Hwang** [2] **Mehak Arora** [1] **Rishikesan Kamaleswaran** [3 4]

Abstract

Knowledge distillation (KD) offers a natural path to compress large biomedical vision-language models into smaller students that meet the compute, latency, and data-locality constraints of clinical deployment. However, KD suffers from *catastrophic out-of-distribution (OOD) forgetting*, where fine-tuning on a narrow mix degrades generalization to unseen modalities and task structures. We introduce *Hard-First*, an entropy-guided curriculum that orders samples from highest to lowest teacher predictive entropy, prioritizing hard examples to better preserve decision-boundary information. To evaluate both transfer and retention, we design a two-axis OOD framework that separates visual-modality shifts (unseen imaging modalities; MedBookVQA) from structural shifts (multi-image reasoning; MedFrameQA). Across 10 configurations on five benchmarks with Qwen2.5-VL-72B→3B, *Hard-First* achieves the largest transfer gain (PathVQA $+8.6$ pp over supervised fine-tuning) while minimizing OOD degradation (MedBookVQA -2.7 pp, MedFrameQA -4.5 pp). Cross-family replication (InternVL3-38B→2B) further establishes *Hard-First* as the top OOD-preserving ordering.

1. Introduction

Recent advances in vision-language models (Chen et al., 2024; Bai et al., 2025) have enabled strong performance on medical imaging tasks (Li et al., 2023a; Moor et al., 2023; Zhang et al., 2023), but these gains largely depend on large-scale models. Deploying such models at the point of care is constrained by hardware availability, data-locality regulations, and strict latency requirements, all of which favor substantially smaller models. Knowledge distillation (KD) (Hinton et al., 2015) offers a practical solution by transferring knowledge from a large teacher to a compact student via soft-label training.

KD for medical imaging presents challenges not typically encountered in generic domains. Medical imaging encompasses diverse modalities (e.g., radiographs, CT, MRI, ultrasound, and pathology), yet practical training datasets cover only a limited subset. While a large teacher can leverage its capacity and broad pre-training to capture this diversity, a student distilled on a narrow subset may lose competence outside that distribution, leading to a form of catastrophic forgetting in which out-of-distribution (OOD) generalization is traded for improved in-distribution performance. Importantly, the teacher's soft-label distribution encodes sample difficulty: high-entropy predictions reflect uncertainty and complex decision boundaries, whereas low-entropy predictions indicate more straightforward cases. This asymmetry implies that samples are not equally informative for distillation, and that their ordering may play a critical role.

We identify three gaps in the existing literature. First, medical VLM evaluations (Li et al., 2023a; Moor et al., 2023; Zhang et al., 2023) typically report per-benchmark accuracy without categorizing the underlying distribution shift. Qualitatively different OOD axes such as novel imaging modalities and multi-image reasoning are therefore conflated, obscuring which capabilities are degraded. Second, prior entropy-aware KD methods (Su et al., 2025) inject $\bar{H}_T(x)$ only as a per-sample weight on the distillation loss, leaving its role as a training-order signal unexplored (Section B). This injection point is especially consequential for low-rank adaptation, where the order of early gradient steps determines how limited adapter capacity is committed. Third, prior curriculum-based KD approaches (Li et al., 2023c; Jafari et al., 2021; Wang et al., 2022) schedule temperature or follow an easy-to-hard progression (Li et al., 2023b); whether this easy-to-hard direction remains optimal in multi-domain VLM distillation under an OOD-preservation objective is unclear.

To address these gaps, we make three contributions:

[1]Department of Electrical and Computer Engineering, Duke University, USA [2]Department of Mathematics, Korea Military Academy, Republic of Korea [3]Department of Surgery, Duke University, USA [4]Department of Anesthesiology, Duke University, USA. Correspondence to: Junseob Kim <junseob.kim@duke.edu>.

Presented at the ICML 2026 Workshop "Continual Adaptation at Scale: Towards Sustainable AI".

1. *Hard→Easy entropy-guided curriculum distillation.* We compute $\bar{H}_T(x)$ offline and sort the training data in descending entropy, applying the signal as a training-order curriculum rather than as a per-sample loss weight or temperature schedule. This is the only configuration that simultaneously achieves the best transfer and the best OOD preservation across 10 configurations on five benchmarks.
2. A *two-axis OOD evaluation framework* for biomedical VLM distillation that disentangles qualitatively distinct shifts. We pair MedBookVQA (Yip et al., 2025) (modality-shift OOD: 40.7% of modalities absent from training) with MedFrameQA (Yu et al., 2025) (structural-shift OOD: multi-image cross-frame reasoning absent from training).
3. A *systematic comparison of three entropy injection points*—loss weighting, temperature, and training order—showing that the injection point, rather than the entropy value itself, determines whether transfer and preservation can coexist.

Together, these contributions highlight the importance of training dynamics for preserving OOD generalization during distillation. Implementation details, baselines, and supporting analyses are in the appendix.

2. Method

Setup. Let teacher T and student S produce token-level conditional distributions $p_T(\cdot \mid x, y_{<t})$ and $p_S(\cdot \mid x, y_{<t})$ over a vocabulary $\mathcal{V}$, where x is a (question, image) input and $y = (y_1, \ldots, y_L)$ is the answer-token sequence. We save the teacher's top-K logits (K=200) per answer token offline; the KL term in the loss is computed by re-expanding this sparse representation to the full vocabulary. The student is trained via LoRA (Hu et al., 2022) adapters on the frozen base.

Per-sample teacher entropy. We define $\bar{H}_T(x)$ as the mean token-level entropy of the top-K teacher distribution over the answer span:

$$\bar{H}_T(x) \;=\; \frac{1}{L}\sum_{t=1}^{L} H(p_T(\cdot \mid x, y_{<t}))\,, \tag{1}$$

where $H(p) := -\sum_v p(v)\log p(v)$ is the Shannon entropy (in nats). Higher $\bar{H}_T(x)$ marks samples where the teacher spreads probability across plausible alternatives, so the soft-label distribution carries information beyond the hard label. Per-dataset statistics are reported in Table 7 (Section D).

What "hard" and "easy" look like. Across the training examples, $\bar{H}_T$ has a mean of 0.88, with the bottom quartile at 0.59 and the top quartile at 1.14. Low-entropy samples ($\bar{H}_T \lesssim 0.5$) are predominantly binary (yes/no) questions where the teacher assigns >95% probability to one answer (e.g., *"Is there pneumothorax on the left side?"*). High-entropy samples ($\bar{H}_T \gtrsim 1.2$) are open-ended or differential-diagnosis questions where the teacher spreads probability across multiple plausible continuations (e.g., *"What is the most likely diagnosis?"*, with 3–4 candidate conditions each receiving 15–30% probability). These boundary-adjacent samples are the ones our curriculum exploits.

Base loss. For our curriculum to exploit teacher uncertainty, the base loss must preserve per-sample uncertainty signal rather than down-weight it—so we adopt *entropy-conditioned temperature* (ECT) over alternatives such as fixed-τ KD (Hinton et al., 2015) or iteration-scheduled τ (Li et al., 2023c; Jafari et al., 2021), both of which apply a uniform τ across the batch. ECT sets τ per sample from the teacher's entropy:

$$\tau(x) = \tau_{\text{base}} + \gamma\,\tilde{H}_T(x),$$

where $\tilde{H}_T(x)$ denotes $\bar{H}_T(x)$ rescaled to lie in $[0,1]$, and $p^{\tau}(\cdot) \propto p(\cdot)^{1/\tau}$. The per-sample loss is

$$\begin{aligned}\mathcal{L}(x) &= (1-\alpha)\,\mathrm{CE}(p_S, y) \\ &\quad + \alpha\,\tau(x)^2 \cdot \frac{1}{L}\sum_{t=1}^{L} \mathrm{KL}\Big(p_T^{\tau(x)} \,\|\, p_S^{\tau(x)}\Big)_t\,,\end{aligned} \tag{2}$$

where the first term is token-level cross-entropy over the answer span and the second term is mean token-level KL divergence (teacher as target). Hyperparameters τ_{base}, γ, α are listed in Section A.

Hard→Easy curriculum. Given training set $\{x_i\}_{i=1}^{N}$, we compute $\bar{H}_T(x_i)$ once and define a permutation $\sigma : [N] \to [N]$ that sorts samples in descending entropy:

$$\bar{H}_T(x_{\sigma(1)}) \geq \bar{H}_T(x_{\sigma(2)}) \geq \cdots \geq \bar{H}_T(x_{\sigma(N)}). \tag{3}$$

We feed batches to the optimizer in the order σ, with `shuffle=False`, so the same entropy-descending order is preserved across all training epochs. The loss (2) and hyperparameters are unchanged; only the ordering of batches differs. This is our proposed injection point: $\bar{H}_T$ as a sample-order signal rather than a loss weight or temperature modulator. Under Hard→Easy, the optimizer sees high-entropy samples first, before the LoRA adapter has committed its limited capacity to a specific decision surface.

3. Experiments

Setup. We use `Qwen2.5-VL-72B-Instruct` as the teacher and `Qwen2.5-VL-3B-Instruct` as the student (Bai et al., 2025). We adopt a general-purpose

Algorithm 1 Hard-First Curriculum Distillation

Input: train set $\{x_i\}_{i=1}^N$, teacher T, LoRA student S_θ, $\tau_{\text{base}}, \gamma, \alpha$, top-$K$
Offline (one-time).
  For each x_i, compute $\bar{H}_i = \bar{H}_T(x_i)$ via teacher top-K logits (Equation (1)); rescale $\tilde{H}_i = \bar{H}_i / \log|\mathcal{V}|$
  $\sigma \leftarrow \text{argsort}(\{\bar{H}_i\}, \text{descending})$ (Hard→Easy)
Training (E **epochs,** `shuffle=False`**).**
for $e = 1$ **to** E **do**
  for i **in** σ (fixed order) **do**
    $\tau_i \leftarrow \tau_{\text{base}} + \gamma \cdot \tilde{H}_i$
    $\mathcal{L} \leftarrow (1-\alpha)\,\text{CE}(p_{S_\theta}, y_i) + \alpha\,\tau_i^2\,\text{KL}\big(p_T^{\tau_i} \,\|\, p_{S_\theta}^{\tau_i}\big)$
    $\theta \leftarrow \theta - \eta\,\nabla_\theta\,\mathcal{L}$ (LoRA only)
  end for
end for

72B teacher rather than a medical-specialized model (e.g., LLaVA-Med (Li et al., 2023a)) for two reasons: the 24× capacity gap provides a strong soft-label signal, and it isolates the effect of the distillation protocol from domain-specific pre-training in the teacher. We train the student using LoRA with AdamW on 13,030 single-image examples from VQA-RAD (Lau et al., 2018), PathVQA (He et al., 2020), and SLAKE (Liu et al., 2021). Evaluation uses constrained decoding on five held-out benchmarks: the three training datasets, along with MedBookVQA (Yip et al., 2025) and MedFrameQA (Yu et al., 2025) as OOD evaluation axes.

Configurations. We group methods by *how the teacher entropy signal is used* during training.

- **No-KD / generic KD (no entropy signal).** SFT (CE only); Standard KD (Hinton et al., 2015); DKD (Zhao et al., 2022); DIST (Huang et al., 2022).
- **Entropy as loss weight.** Three baseline variants we implement: entropy-gated KL (per-sample KL weight decays with $\bar{H}_T$); teacher–student gap weighting; adaptive α (sigmoid of teacher and student entropies to re-balance CE vs. KL).
- **Entropy as temperature (± training order).** Using ECT (Equation (2)) as the base loss, we compare three sample orderings: random shuffle (no curriculum), Easy→Hard, and Hard→Easy. All three share the same loss, data, and optimizer steps; they differ only in the ordering of batches.

Findings. (i) *The injection point of the entropy signal determines performance* (Figure 1). Loss-weighting variants improve PathVQA by +0.8 to +5.4 pp but show no consistent OOD gains (MedBook remains within ±0.6 pp of SFT; MedFrame drops by 2–4 pp). ECT with random shuffling yields +5.1 pp on PathVQA yet leaves MedBook near SFT (−0.2 pp) and degrades MedFrame (−3.0 pp). In contrast, Hard→Easy under the same ECT loss improves

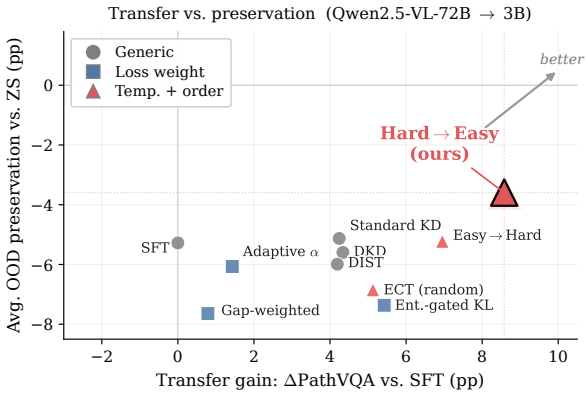

Figure 1. Transfer (ΔPathVQA vs. SFT) vs. OOD preservation (mean(ΔMedBook, ΔMedFrame) vs. ZS student) across 10 configurations. Higher is better on both axes; Hard→Easy with ECT achieves the highest transfer and the best OOD preservation.

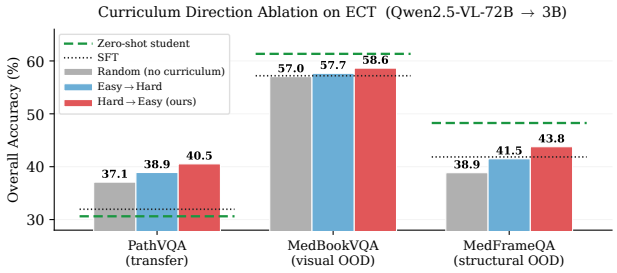

Figure 2. Curriculum direction ablation on ECT: the three configurations differ only in batch ordering. All metrics improve monotonically Random < Easy→Hard < Hard→Easy. Green dashed: ZS student; black dotted: SFT.

PathVQA by +8.6 pp while narrowing both OOD gaps, uniquely among all 10 configurations in Table 1. (ii) *Curriculum direction is not symmetric* (Figure 2). Across four of five benchmarks (PathVQA, SLAKE, MedBookVQA, MedFrameQA), every metric improves monotonically Random < Easy→Hard < Hard→Easy; VQA-RAD differences fall within a 2.5 pp noise band. This contrasts with the easy→hard schedule of Li et al. (2023b) for data-free CNN distillation; we attribute the divergence to setting (Section 4). (iii) *The preservation effect is architecture-invariant.* On a second family (InternVL3-38B→2B), Hard→Easy remains the top OOD-preserving ordering; the transfer gain collapses to within noise under the smaller capacity gap (Section G).

4. Discussion

Where entropy is injected matters more than the entropy value itself. All entropy-aware methods in Table 1—loss-weight variants, ECT with random shuffling, and our Hard→Easy—share the same $\bar{H}_T(x)$ signal. Yet only Hard→Easy consistently outperforms all baselines on MedBook and MedFrame. Loss-weight and temperature-based approaches modulate the per-step gradient magnitude,

Table 1. Overall accuracy (%) on five benchmarks (Qwen-VL-72B→3B). The three in-distribution benchmarks are trained on; MedBookVQA tests visual-modality OOD and MedFrameQA tests structural OOD. Shaded row marks our proposed method, which achieves the best PathVQA, MedBook, and MedFrame and ties for the best SLAKE. Per-answer-type breakdown for PathVQA in Table 6.

	IN-DISTRIBUTION			OUT-OF-DISTRIBUTION	
Method	VQA-RAD	PathVQA	SLAKE	MedBook	MedFrame
Zero-shot reference					
Student 3B	52.77	30.62	51.93	61.34	48.26
Teacher 72B	66.08	32.89	58.62	71.82	47.63
No-KD / generic KD					
SFT	52.33	31.95	57.96	57.20	41.84
Standard KD (Hinton et al., 2015)	53.88	36.19	57.78	57.08	42.27
DKD (Zhao et al., 2022)	52.55	36.29	57.87	57.72	40.69
DIST (Huang et al., 2022)	54.32	36.14	58.62	56.62	41.00
Entropy as loss weight					
Entropy-gated KL	54.55	37.38	56.08	56.92	37.95
Gap-weighted KL	53.66	32.74	58.34	56.82	37.50
Adaptive α	55.43	33.38	57.96	57.76	39.71
Entropy as temperature + training order					
ECT, random shuffle	**56.10**	37.08	58.25	57.04	38.86
ECT, Easy→Hard	53.66	38.91	58.62	57.66	41.49
ECT, Hard→Easy (ours)	55.43	**40.53**	**58.62**	**58.64**	**43.77**

whereas training-order injection determines *when* informative (high-entropy) updates occur relative to the adapter's early capacity allocation. This temporal placement is what drives OOD preservation; in contrast, *selection* alone (top-50% highest-entropy) underperforms random shuffle (Section F).

Why Hard→Easy preserves. Hard→Easy front-loads samples with the largest teacher CE gradient (r=0.40 with $\bar{H}_T$; Section E). Fitting low-entropy samples first commits the adapter to decision boundaries tailored to easy cases, making it difficult for later high-entropy samples to reshape the boundary without disrupting earlier fits. In contrast, fitting high-entropy samples first establishes decision boundaries based on informative signals; subsequent low-entropy samples are already correctly classified and require minimal adjustment. The progression Random < Easy→Hard < Hard→Easy is monotonic on the four information-bearing axes (Figure 2), consistent with the LIFO forgetting pattern (Hacohen & Tuytelaars, 2024), where examples learned earlier are retained longer.

Why hard-first works in this setting. The easy-to-hard finding of Li et al. (2023b) arises from from-scratch CNN students trained on single-domain data, whereas our setting differs on both axes. The student is pre-trained, shifting the objective from representation learning to adaptation, and the training data spans multiple medical domains, where easy samples are dominated by domain-specific binary responses. An easy-to-hard schedule thus risks overcommitting limited adapter capacity to a single domain. In contrast, hard samples lie closer to the base model's decision surface and induce cross-domain boundaries that easy-first fails to capture. Cross-family replication on InternVL3 (Section G) suggests that this mechanism is not architecture-specific, consistent with prior hard-first regimes (Shrivastava et al., 2016; Jarca et al., 2025).

Limitations. Our 13 k-example, 3-epoch LoRA adaptation budget is consistent with standard medical VLM fine-tuning practice; PathVQA OPEN annotation quirks and further scope caveats are discussed in Section H.

5. Conclusion

Biomedical vision-language distillation exhibits a measurable trade-off between transfer and preservation, which becomes apparent only when OOD is decomposed into visual and structural axes. Across 10 configurations, only Hard→Easy, using entropy as a training-order signal, achieves both the best transfer and the strongest OOD preservation. Cross-family replication confirms preservation as architecture-invariant. What matters is not the entropy signal itself, but where it is injected.

Impact Statement

This work studies KD for medical VLMs. Smaller distilled models can lower computational barriers to deployment, improving accessibility in resource-constrained settings, but may also enable premature or unvetted clinical use. We do not endorse deployment without appropriate regulatory review and clinical validation. All experiments use public datasets and involve no patient-identifying information.

References

Bai, S. et al. Qwen2.5-VL technical report. *arXiv preprint arXiv:2502.13923*, 2025.

Bengio, Y., Louradour, J., Collobert, R., and Weston, J. Curriculum learning. In *International Conference on Machine Learning (ICML)*, 2009.

Chen, Z., Wu, J., Wang, W., Su, W., Chen, G., Xing, S., Zhong, M., Zhang, Q., Zhu, X., Lu, L., et al. InternVL: Scaling up vision foundation models and aligning for generic visual-linguistic tasks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

French, R. M. Catastrophic forgetting in connectionist networks. *Trends in Cognitive Sciences*, 3(4):128–135, 1999.

Hacohen, G. and Tuytelaars, T. Forgetting order of continual learning: Examples that are learned first are forgotten last. *arXiv preprint arXiv:2406.09935*, 2024.

Hacohen, G. and Weinshall, D. On the power of curriculum learning in training deep networks. In *International Conference on Machine Learning (ICML)*, 2019.

He, X., Zhang, Y., Mou, L., Xing, E., and Xie, P. PathVQA: 30000+ questions for medical visual question answering. *arXiv preprint arXiv:2003.10286*, 2020.

Hinton, G., Vinyals, O., and Dean, J. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., and Chen, W. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations (ICLR)*, 2022.

Huang, T., You, S., Wang, F., Qian, C., and Xu, C. Knowledge distillation from a stronger teacher. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

Jafari, A., Rezagholizadeh, M., Sharma, P., and Ghodsi, A. Annealing knowledge distillation. In *Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, 2021.

Jarca, A., Croitoru, F.-A., and Ionescu, R. T. Task-informed anti-curriculum by masking improves downstream performance on text. In *Findings of the Association for Computational Linguistics (ACL)*, 2025. arXiv:2502.12953.

Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, 114(13):3521–3526, 2017.

Kumar, M. P., Packer, B., and Koller, D. Self-paced learning for latent variable models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2010.

Lau, J. J., Gayen, S., Ben Abacha, A., and Demner-Fushman, D. A dataset of clinically generated visual questions and answers about radiology images. *Scientific Data*, 5(1):180251, 2018.

Li, C., Wong, C., Zhang, S., Usuyama, N., Liu, H., Yang, J., Naumann, T., Poon, H., and Gao, J. LLaVA-Med: Training a large language-and-vision assistant for biomedicine in one day. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023a.

Li, J., Zhou, S., Li, L., Wang, H., Yu, Z., and Bu, J. Dynamic data-free knowledge distillation by easy-to-hard learning strategy. *Information Sciences*, 642, 2023b. arXiv:2208.13648.

Li, Z., Li, X., Yang, L., Zhao, B., Song, R., Luo, L., Li, J., and Yang, J. Curriculum temperature for knowledge distillation. In *AAAI Conference on Artificial Intelligence*, 2023c.

Liu, B., Zhan, L.-M., Xu, L., Ma, L., Yang, Y., and Wu, X.-M. SLAKE: A semantically-labeled knowledge-enhanced dataset for medical visual question answering. In *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2021.

McCloskey, M. and Cohen, N. J. Catastrophic interference in connectionist networks: The sequential learning problem. *Psychology of Learning and Motivation*, 24: 109–165, 1989.

Moor, M., Huang, Q., Wu, S., Yasunaga, M., Dalmia, Y., Leskovec, J., Zakka, C., Reis, E. P., and Rajpurkar, P. Med-Flamingo: a multimodal medical few-shot learner. In *Machine Learning for Health (ML4H)*, 2023.

Shrivastava, A., Gupta, A., and Girshick, R. Training region-based object detectors with online hard example mining. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

Su, C.-P., Tseng, C.-H., Pu, B., Zhao, L., Yang, J., Chen, Z., and Lee, S.-J. EA-KD: Entropy-based adaptive knowledge distillation. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2025. arXiv:2311.13621.

Toneva, M., Sordoni, A., Tachet des Combes, R., Trischler, A., Bengio, Y., and Gordon, G. J. An empirical study of example forgetting during deep neural network learning. In *International Conference on Learning Representations (ICLR)*, 2019.

Wang, C., Yang, K., Zhang, S., Huang, G., and Song, S. TC3KD: Knowledge distillation via teacher-student cooperative curriculum customization. *Neurocomputing*, 508: 284–292, 2022. doi: 10.1016/j.neucom.2022.07.055.

Wu, X., Dyer, E., and Neyshabur, B. When do curricula work? In *International Conference on Learning Representations (ICLR)*, 2021.

Yip, S. L., He, S., Nie, Y., Chan, S. P., Ye, Y., Lam, S. Y., and Chen, H. MedBookVQA: A systematic and comprehensive medical benchmark derived from open-access book. *arXiv preprint arXiv:2506.00855*, 2025.

Yu, S., Wang, H., Wu, J., Luo, L., Wang, J., Xie, C., Rajpurkar, P., Yang, C., Yang, Y., Wang, K., Yu, Y., and Zhou, Y. MedFrameQA: A multi-image medical VQA benchmark for clinical reasoning. *arXiv preprint arXiv:2505.16964*, 2025.

Zhang, K., Yu, J., Yan, Z., Liu, Y., Adhikarla, E., Fu, S., Chen, X., Chen, C., Zhou, Y., Li, X., He, L., Davison, B. D., Li, Q., Chen, Y., Liu, H., and Sun, L. BiomedGPT: A unified and generalist biomedical generative pre-trained transformer for vision, language, and multimodal tasks. *arXiv preprint arXiv:2305.17100*, 2023.

Zhao, B., Cui, Q., Song, R., Qiu, Y., and Liang, J. Decoupled knowledge distillation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

A. Implementation Details

Figure 3 visualises the full pipeline at a glance: (a) what teacher entropy $\bar{H}_T(x)$ looks like on real training samples from the three datasets, (b) the offline → online KD flow with the ECT loss and per-sample temperature, and (c) the two-axis OOD evaluation. The remainder of this section spells out the implementation of each component.

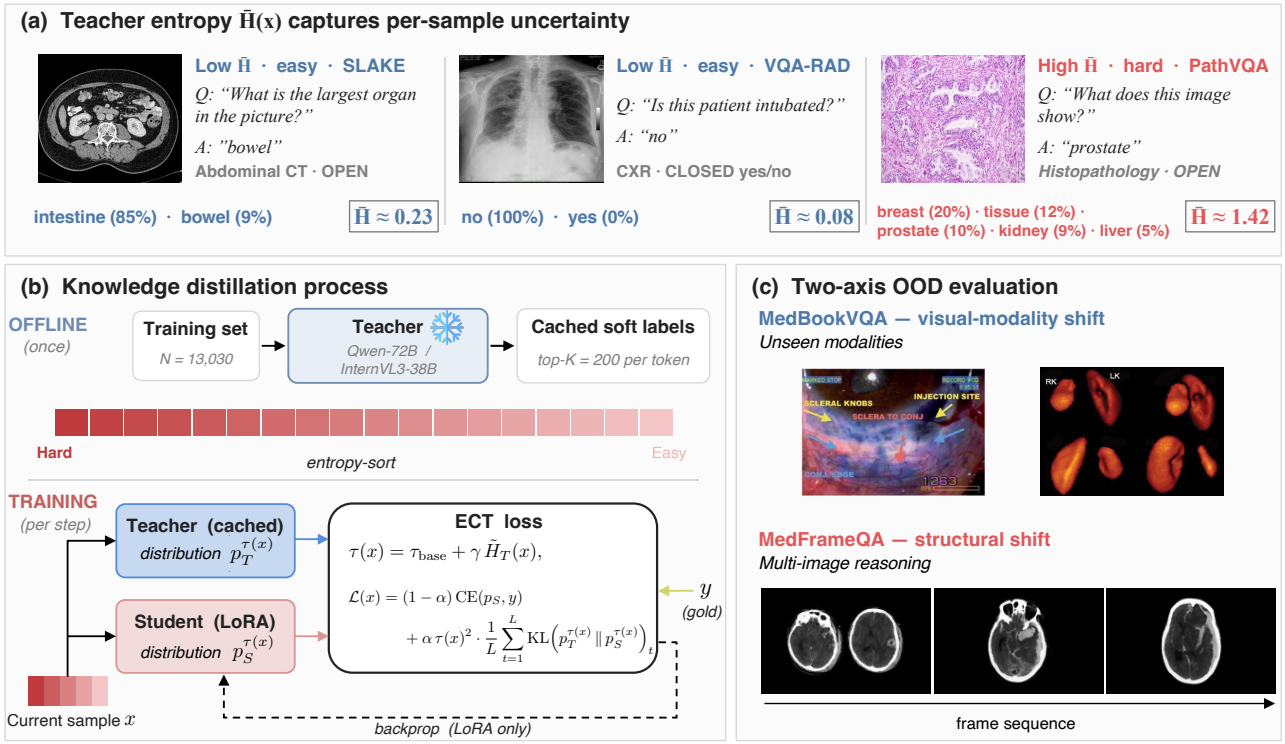

Figure 3. **Method overview. (a)** Teacher entropy $\bar{H}_T(x)$ captures per-sample uncertainty across the multi-domain training mix; percentages next to each sample are the actual cached teacher top-K probabilities at the decisive answer token. $\bar{H}$ ranges from 0.08 (VQA-RAD yes/no, sharp) to 1.42 (PathVQA histology, a 5-way differential over organ of origin where the gold "prostate" is only the teacher's 3rd pick). **(b)** Pipeline. *Offline:* the frozen teacher is run once on the 13,030 training samples; top-K=200 soft labels and $\bar{H}_T(x)$ are cached and samples are sorted in descending $\bar{H}_T$. *Training:* the student (LoRA, r=16) matches the cached teacher distribution via the ECT loss with per-sample temperature $\tau(x)=\tau_{\text{base}}+\gamma\tilde{H}_T(x)$ ($\tilde{H}_T$ rescaled to $[0,1]$); gradients update only the adapter. **(c)** Two-axis OOD evaluation: visual-modality shift (MedBookVQA; 40.7% modalities absent from training) and structural shift (MedFrameQA; 2–5-frame reasoning).

Soft label extraction. Top-200 teacher log-probabilities over answer tokens are extracted offline and stored as sparse tensors, one file per training sample (13,030 files total). During training, these sparse representations are reconstituted to the full vocabulary when computing the KL term.

Per-sample entropy and its rescaling. $\bar{H}_T(x)$ is defined in Equation (1); values are computed offline from the cached top-K teacher distributions, stored in `train_entropy.json`, and read by the DataLoader at training startup for curriculum ordering. For ECT's temperature modulation (Equation (2)), we use the rescaled quantity $\tilde{H}_T(x) = \bar{H}_T(x)/\log|\mathcal{V}|$, which bounds $\tau(x) \in [\tau_{\text{base}}, \tau_{\text{base}}+\gamma]$. The rescaling is strictly monotone, so it does not affect the sort order used for curriculum construction. The full procedure is summarised in Algorithm 1 (main body).

A.1. Baseline Formulations

All baselines use the same overall structure as Equation (2): $\mathcal{L}(x) = (1-\alpha)\,\text{CE}(p_S, y) + \alpha \cdot \mathcal{L}_{\text{KD}}(x)$. Methods differ in $\mathcal{L}_{\text{KD}}$. Notation follows Section 2; $\bar{H}_S(x)$ denotes the student analogue of $\bar{H}_T(x)$, computed on-the-fly using the same normalization. The headings below match the row order and names in Table 1; *entropy-gated KL*, *gap-weighted KL*, and *adaptive* α are baseline variants we implement, while the others are prior work (citations inline).

SFT. No KD term: $\mathcal{L}(x) = \mathrm{CE}(p_S, y)$. No-KD baseline.

Standard KD (Hinton et al., 2015). Temperature-softened KL over the full vocabulary (teacher as target):

$$\mathcal{L}_{\mathrm{KD}}(x) = \tau^2 \cdot \mathrm{KL}(p_T^\tau \,\|\, p_S^\tau)\,.$$

DKD (Zhao et al., 2022). Splits the KL into target-class (TCKD) and non-target-class (NCKD) components. Let $c \in \mathcal{V}$ denote the gold token at position t. TCKD collapses the vocabulary to a binary (target, non-target) distribution and computes KL on that; NCKD renormalizes the mass on $\mathcal{V} \setminus \{c\}$ and computes KL there:

$$\mathcal{L}_{\mathrm{KD}}(x) = \tau^2\big(\alpha_{\mathrm{tc}} \cdot \mathrm{TCKD} + \beta_{\mathrm{nc}} \cdot \mathrm{NCKD}\big),$$

with $\alpha_{\mathrm{tc}}, \beta_{\mathrm{nc}}$ set to the defaults of Zhao et al. (2022).

DIST (Huang et al., 2022). Replaces KL with a Pearson-correlation objective, which is scale-invariant and better suited to a large capacity gap. Let $\rho(\cdot,\cdot)$ denote the mean Pearson correlation between matching rows of two matrices. Writing P_S, P_T for the stacked per-token distributions at temperature τ, DIST combines inter-class (across vocabulary, row-wise) and intra-class (across tokens, column-wise) correlations:

$$\mathcal{L}_{\mathrm{KD}}(x) = \beta_{\mathrm{inter}}\big(1 - \rho(P_S, P_T)\big) + \beta_{\mathrm{intra}}\big(1 - \rho(P_S^\top, P_T^\top)\big).$$

Entropy-gated KL. Multiplies the KL by a per-sample gate derived from teacher entropy:

$$\lambda(x) = \exp\big(-\beta\, \bar{H}_T(x)\big), \qquad \mathcal{L}_{\mathrm{KD}}(x) = \lambda(x) \cdot \tau^2 \cdot \mathrm{KL}(p_T^\tau \,\|\, p_S^\tau)\,.$$

Direction is the opposite of EA-KD (Su et al., 2025): high teacher entropy is treated as a low-trust signal, so the KL contribution decays as $\bar{H}_T$ grows.

Gap-weighted KL. Extends entropy gating with a student-uncertainty factor, distilling only when the teacher is confident and the student still has headroom:

$$\lambda_{\mathrm{gap}}(x) = \exp\big(-\beta\, \bar{H}_T(x)\big) \cdot \big(1 - \exp(-\delta\, \bar{H}_S(x))\big), \quad \mathcal{L}_{\mathrm{KD}}(x) = \lambda_{\mathrm{gap}}(x) \cdot \tau^2 \cdot \mathrm{KL}(p_T^\tau \,\|\, p_S^\tau)\,.$$

Adaptive α. Replaces static CE/KL weighting with a per-sample sigmoid and combines it with ECT's per-sample temperature. CE weight $(1 - \alpha(x))$ and KL weight $\alpha(x)$ sum to 1, so reducing KL automatically redistributes supervision back to CE:

$$\begin{aligned}
\tau(x) &= \tau_{\mathrm{base}} + \gamma\, \tilde{H}_T(x), \\
\alpha(x) &= \sigma\big(w_t(1 - \bar{H}_T(x)) + w_s \bar{H}_S(x) - b\big), \\
\mathcal{L}(x) &= \big(1 - \alpha(x)\big)\, \mathrm{CE}(p_S, y) + \alpha(x)\, \tau(x)^2\, \mathrm{KL}\Big(p_T^{\tau(x)} \,\|\, p_S^{\tau(x)}\Big)\,.
\end{aligned}$$

$\alpha(x)$ is high when the teacher is confident and the student is uncertain.

ECT: random shuffle, Easy→Hard, Hard→Easy (ours). The shared base loss is defined in Equation (2). The three rows of Table 1 for ECT use identical loss, data, optimizer steps, and hyperparameters; they differ only in the permutation σ over training samples (Equation (3)). Random shuffle uses σ drawn uniformly and reshuffled every epoch; Easy→Hard fixes σ in ascending $\bar{H}_T$; Hard→Easy (our proposed method) fixes σ in descending $\bar{H}_T$.

Training hyperparameters.

LoRA	rank 16, α=32, dropout 0.05; targets $\{q, k, v, o\}$
Optimizer	AdamW; lr 2×10^{-4}; warmup 0.03 (linear); grad clip 1.0
Training	3 epochs, effective batch 32 (1×32 grad-accum), bf16, max seq 1024
KD (shared)	τ_{base}=4, γ=1, α_{KD}=0.5, top-K=200

Method-specific KD hyperparameters are listed in Section A.1. Each training run takes ∼3.5 h on a single H200.

Evaluation. Constrained decoding (format hints for yes/no and MCQ letter), `max_new_tokens`=32, batch size 4. A full evaluation across the five benchmarks takes ∼3 h on a single GPU.

B. Related Work

Knowledge distillation. Since Hinton et al. (2015), a line of work has improved the soft-label KL objective: Zhao et al. (2022) decouple target-class and non-target KL, Huang et al. (2022) use Pearson correlation. Several authors adapt temperature or loss weight per sample based on teacher entropy. Our base loss (ECT, Equation (2)) falls in this family; the novelty we claim is not ECT itself but the training-order curriculum added on top.

Curriculum + KD and its direction. Several works combine curriculum with KD but target different axes. CTKD (Li et al., 2023c) and Annealing-KD (Jafari et al., 2021) schedule the distillation temperature, not sample order. TC3KD (Wang et al., 2022) and CuDFKD (Li et al., 2023b) order samples by teacher- or student-estimated difficulty; Li et al. (2023b) in particular adopts an easy→hard schedule for data-free CNN distillation. Entropy has also been used as a KD loss signal: EA-KD (Su et al., 2025) re-weights samples by entropy but does not reorder training. Our contribution is orthogonal: offline teacher predictive entropy as a one-shot descending sort key on training order (Hard→Easy), not a loss weight or temperature schedule. The easy→hard choice in Li et al. (2023b) motivates us to control setting: we operate on large-teacher VLM LoRA adaptation with multi-domain biomedical data and OOD-forgetting as the primary metric, where the learnability-driven easy-first intuition no longer governs.

Curriculum direction in other settings. Whether easy→hard or hard→easy wins is setting-dependent. Bengio et al. (2009) and self-paced learning (Kumar et al., 2010) favor easy-first as a warm start; Hacohen & Weinshall (2019) confirm a moderate easy-first benefit; Wu et al. (2021) argue curricula help primarily under limited budgets or noisy labels. Against this, hard-first strategies dominate in several regimes: OHEM (Shrivastava et al., 2016) for object detection and TIACBM (Jarca et al., 2025) for masked-language-model pretraining. Our finding is consistent with this broader pattern and, to our knowledge, is the first demonstration in a KD setting.

Example-forgetting. Toneva et al. (2019) formalize per-example forgetting events in SGD training; Hacohen & Tuytelaars (2024) observe a LIFO pattern — examples learned first are forgotten last. This directly motivates Hard→Easy: committing to hard samples early means they are reinforced rather than overwritten. Classical forgetting work (McCloskey & Cohen, 1989; French, 1999) and mitigations such as EWC (Kirkpatrick et al., 2017) address forgetting across tasks; we treat single-task forgetting of base-model OOD capability during distillation.

C. Dataset Design and Two-Axis OOD Framework

This appendix details the training data, held-in evaluation, and the construction and modality-level verification of the two OOD test sets. The design supports two claims a distillation study must answer separately: does KD *transfer* teacher knowledge to related unseen tasks, and does KD *preserve* the student's pre-existing capabilities on inputs outside the training distribution?

C.1. Training Set

The training mix combines three public medical VQA datasets (Table 2).

Table 2. Training set: 13,030 single-image VQA samples across three public medical datasets.

Dataset	N	Format	Modality
VQA-RAD (Lau et al., 2018)	1,793	CLOSED (58%) + OPEN (42%)	CT / X-ray / MRI
PathVQA (He et al., 2020)	9,635	CLOSED (52%) + OPEN (48%)	Pathology textbook
SLAKE (Liu et al., 2021)	1,602	CLOSED (42%) + OPEN (58%)	Multi-organ radiology
Total	13,030		

All training samples are single-image. Both CLOSED and OPEN answer formats are represented so the student is not biased toward a single prompt structure. In-distribution evaluation uses the official test splits released by the dataset authors without

any re-splitting: VQA-RAD (451), PathVQA (2,028), SLAKE (1,061). The SLAKE test split is relatively large because its official English release (`BoKelvin/SLAKE` + `mdwiratathya/SLAKE-vqa-english`) uses an approximately 60/20/20 train/val/test partition, while VQA-RAD and PathVQA were released with smaller test partitions.

C.2. Out-of-Distribution Benchmarks

We evaluate on two held-out datasets chosen to stress different aspects of generalization (Table 3).

Table 3. Two-axis OOD evaluation: MedBookVQA probes visual-modality shift and MedFrameQA probes structural (multi-image) shift.

Dataset	N	Format	Modality
MedBookVQA (Yip et al., 2025)	5,000	MCQ (4 options), single-image	18 modalities
MedFrameQA (Yu et al., 2025)	2,851	MCQ (5–9 options), multi-image (2–5 frames)	CT / MRI / US / X-ray

Visual OOD — MedBookVQA. Single-image 4-option MCQ, matching the training answer format. We cross-check the per-modality composition against the training set (Table 4).

Table 4. MedBookVQA modality overlap with the training set: 55.2% present, 4.1% partial, 40.7% absent.

Training presence	Modalities	N (%)
In training	CT (664), MRI (658), Radiography (550), Histopathology (480), Fundus (193), OCT (160), Microscopy (55)	2,760 (55.2%)
Partial overlap	Cardiovascular Imaging (178), Ophthalmic Imaging (28)	206 (4.1%)
Absent from training	General Photo of Affected Area (853), Visible Light Photography (551), Nuclear Imaging (510), Elastography (20), EIT (11), Printed Signal waves (22), Infrared Reflectance (5), Microperimetry (1), Other (61)	**2,034 (40.7%)**

The absent portion is dominated by regular clinical photographs of body parts ("bandaged arm", "lower extremity affected area"), visible-light photography, PET/SPECT, and small quantities of specialty modalities (Elastography, EIT, Infrared Reflectance).

Structural OOD — MedFrameQA. Each question comes with 2–5 frames and requires cross-frame reasoning. Modality breakdown: CT 38%, MRI 34%, ultrasound 14%, X-ray 11%, other 5% — largely overlapping with the training modalities (ultrasound is novel; pathology is absent). The primary shift is structural: multi-image inputs with 5–9 answer options, a task structure not present in training.

C.3. Why Both Axes

Zero-shot (ZS) teacher–student gap differs sharply between the two (Table 5).

Table 5. Zero-shot teacher–student gap on the two OOD axes: transfer vs. preservation regime.

	Teacher 72B	Student 3B (ZS)	Gap	Regime
MedBookVQA	71.82	61.34	$+10.5$	teacher headroom (transfer axis)
MedFrameQA	47.63	48.26	-0.7	teacher $\approx$ student (preservation axis)

MedBook probes whether distillation can *transfer* the teacher's advantage on a related but unseen task; MedFrame probes whether distillation *preserves* the student's pre-existing multi-image capability. Empirically (Table 1) every KD configuration degrades both axes relative to ZS, so the two-axis split is read as preservation on each axis, and a method that loses more on one than on the other reveals which kind of capability KD is destroying.

C.4. Known Dataset Issues

- PathVQA OPEN: $\sim$46% of OPEN items use single-token classification labels as gold answers, yielding token-F1 of 2.5–4.5% across *all* methods including the teacher. Not a method failure. Table 6 breaks down PathVQA by answer

type: OVERALL gains concentrate in the CLOSED (yes/no) subset, which is the reliable signal; OPEN F1 is near-floor for every method.

- Teacher PathVQA (32.89) $<$ distilled students: the zero-shot 72B teacher never sees the PathVQA training split, so it does not learn PathVQA's single-token OPEN annotation convention and scores OPEN F1 $= 4.4$ (near floor). Fine-tuned students inherit this convention from three epochs on the training set, so OVERALL can surpass the teacher even though the teacher provides the soft-label supervision. This is a standard student-surpasses-teacher effect in transfer learning on tasks with distinctive label conventions and is not evidence of distillation failure.

- MedBookVQA small modality labels: a small number of items have ambiguous modality tags (e.g., EIT labelled mammogram questions); these are rare and do not affect the 40.7% novel-modality tally materially.

Table 6. PathVQA answer-type breakdown. The yes/no CLOSED subset is the reliable portion of PathVQA; OPEN F1 is depressed near a common floor by annotation noise rather than by method differences. Hard→Easy's CLOSED accuracy of 82.24 is $+16.82$ pp above SFT (the largest CLOSED gap among all methods); the corresponding OVERALL gain reported in the abstract is $+8.6$ pp.

Method	CLOSED (%)	OPEN F1 (%)
Student 3B (ZS)	61.16	3.73
SFT	65.42	3.14
Standard KD	74.25	3.27
Entropy-gated KL	76.32	3.49
Gap-weighted KL	67.91	2.50
Adaptive α	69.06	2.53
ECT, Random order	76.12	2.78
ECT, Easy→Hard	79.85	2.94
ECT, Hard→Easy	**82.24**	**3.74**

D. Entropy Signal and Curriculum Composition

This appendix examines the entropy signal itself—its distribution across the training mix and its effect on the composition of batches under Hard→Easy ordering—and rules out the most natural confounder of our main result.

D.1. Per-dataset entropy distribution

Table 7. Per-sample teacher entropy $\bar{H}_T$ statistics (in nats, computed on the renormalized top-K=200 teacher distribution) for the three training datasets on which our curriculum is applied.

Dataset	N	Format	Mean	Q_1	Median	Q_3
VQA-RAD (train)	1,793	CLOSED + OPEN	0.75	0.52	0.78	0.97
PathVQA (train)	9,635	CLOSED + OPEN	**0.94**	0.64	0.98	1.20
SLAKE (train)	1,602	CLOSED + OPEN	0.67	0.49	0.67	0.85

Figure 4 (left) shows the per-dataset entropy histograms. PathVQA has the highest mean entropy (0.94) because its open-ended questions ("what is present?", differential diagnoses) spread teacher probability across multiple plausible tokens. SLAKE has the lowest (0.67), reflecting simpler anatomical questions. By answer type, PathVQA OPEN items average $\bar{H}_T = 1.11$ whereas PathVQA CLOSED average 0.77; VQA-RAD and SLAKE show no systematic CLOSED/OPEN split.

D.2. Curriculum composition under Hard→Easy

Because PathVQA dominates high-entropy samples, Hard→Easy front-loads PathVQA disproportionately (Figure 4 d). Quantitatively, PathVQA's share of batches reaches 97.5% in the top 5% of the Hard→Easy order (vs. its 73.9% base rate), 96.1% in the top 10%, and 94.1% in the top 20%. Under Random shuffle (panel b) the share stays at the base rate, and under Easy→Hard (panel c) the ordering is reversed, giving VQA-RAD and SLAKE relatively more early exposure.

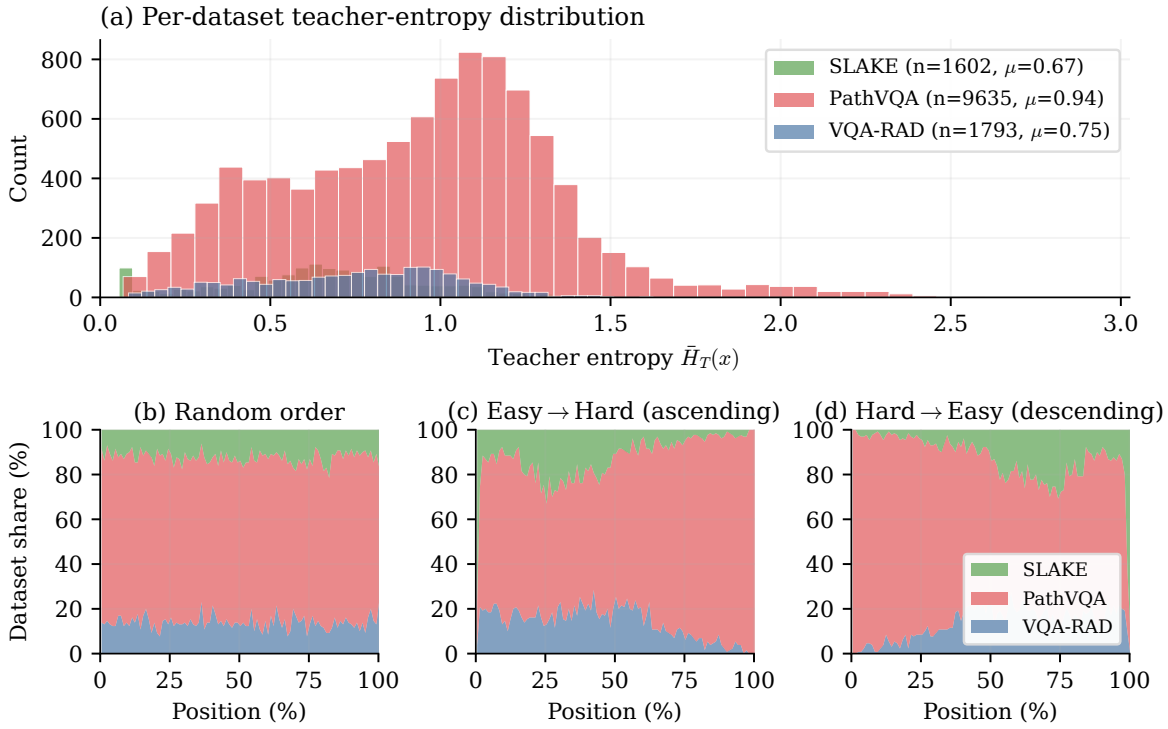

Figure 4. **(a)** Per-dataset teacher entropy distribution on the 13,030-example training set. PathVQA is highest-entropy on average, SLAKE lowest. **(b–d)** Dataset composition along three orderings, as a function of training position (0% = first batch, 100% = last batch). Under Random (b), all three datasets stay at their base rates throughout. Under Easy→Hard (c), low-entropy positions come first, giving VQA-RAD and SLAKE a mild early over-representation. Under Hard→Easy (d), PathVQA dominates the early (high-entropy) positions far beyond its 73.9% base rate, reaching ∼97% in the top 5%.

D.3. Ruling out the composition confounder

A natural concern is that the gain simply reflects PathVQA over-exposure: if Hard→Easy is essentially "PathVQA pretraining followed by brief exposure to other datasets," then perhaps the student acquires general medical-VQA skills that incidentally transfer to every benchmark, including OOD. This would make our method a data-specific artifact rather than a principled curriculum effect.

If that hypothesis were correct, four testable predictions follow from the composition pattern in Figure 4. We show that all four fail.

(P1) Easy→Hard should boost VQA-RAD and SLAKE. Under Easy→Hard, early batches are low-entropy and VQA-RAD/SLAKE are *over*-represented relative to base rate (16.4% and 28.3% in the first 5%, vs. $< 2\%$ under Hard→Easy). The composition hypothesis predicts Easy→Hard should therefore achieve the best VQA-RAD and SLAKE scores. In fact, Easy→Hard reaches VQA-RAD 53.66—the *lowest* among Random (56.10), Easy→Hard (53.66), and Hard→Easy (55.43)—and SLAKE 58.62, identical to Hard→Easy. Front-loading a dataset does not, on its own, improve that dataset.

(P2) Within fixed PathVQA composition, ordering should not matter. All three curricula share identical 73.9% PathVQA composition. If only *exposure amount* mattered, all three should achieve equal PathVQA accuracy. Instead, PathVQA accuracy varies monotonically with ordering: Random 37.08, Easy→Hard 38.91, Hard→Easy 40.53. The +3.45 pp Random→Hard gap exists at fixed composition and must therefore come from ordering *within* the PathVQA samples (hardest PathVQA first vs. easy first).

(P3) The same ordering-only gain appears on OOD. Applying the fixed-composition logic to MedFrameQA (Table 8).

MedFrameQA shows a +4.91 pp gain from ordering alone, larger than the PathVQA gain—the opposite of what "PathVQA-specialization" would predict.

Table 8. Ordering-only gain under ECT at fixed PathVQA composition: Hard→Easy improves all three axes simultaneously.

Ordering	PathVQA	MedBook	MedFrame	*all three gains co-occur?*
Random	37.08	57.04	38.86	—
Easy→Hard	38.91	57.66	41.49	—
Hard→Easy	**40.53**	**58.64**	**43.77**	**Yes**
Ordering-only gain (Hard − Random)	+3.45	+1.60	+4.91	—

(P4) MedFrameQA contains no pathology. Per Section C, MedFrameQA is radiology-only (CT/MRI/ultrasound/X-ray) with zero pathology or histopathology content. Any hypothesis that Hard→Easy transfers "PathVQA-specialized pathology skills" to OOD is therefore falsifiable on MedFrameQA: such transfer has no visual substrate. Yet MedFrameQA improves by $+4.91$ pp over Random. Whatever drives this cannot be PathVQA-specific knowledge.

Taken together, these four observations rule out composition or data-specific pretraining as the causal mechanism. The remaining free variable is the *ordering of samples within their entropy rank*, which isolates the curriculum effect we claim: the adapter commits its capacity to the most boundary-carrying signal first, regardless of which dataset those samples happen to come from.

D.4. Is entropy just a proxy for answer format?

A related concern is whether $\bar{H}_T$ merely tracks the answer-format distribution (CLOSED yes/no vs. OPEN free-form) rather than any notion of intrinsic sample difficulty. If so, Hard→Easy would reduce to "OPEN samples first" and our mechanism claim would be vacuous. The data rules this out.

VQA-RAD inverts the expected relation. Per-format means are: VQA-RAD CLOSED 0.78 vs. OPEN 0.70; PathVQA CLOSED 0.77 vs. OPEN 1.11; SLAKE CLOSED 0.66 vs. OPEN 0.67. If format strictly determined entropy, CLOSED (with a two-element support) should always be lower-entropy than OPEN. VQA-RAD's inversion—where the teacher is *less* confident on yes/no radiology questions than on factoid OPEN questions—shows that entropy reflects genuine per-sample uncertainty, not just format.

Within-format entropy is broadly distributed. Within PathVQA CLOSED alone, $\bar{H}_T$ still ranges from below 0.3 to above 1.5. An ordering by entropy therefore induces a different sequence than an ordering by format, and the $+1.62$ pp PathVQA accuracy gap between Hard→Easy and Easy→Hard (Table 1; see also Section D.3) is produced by this within-format variation specifically—ordering PathVQA samples by entropy, not by format.

The mechanism is format-agnostic. Our claim is that samples where the teacher is uncertain carry richer KL signal, because the softmax distribution contains information about which alternatives the teacher considers close. This holds for OPEN items (many candidate continuations) and CLOSED items (probabilities split between "yes" and "no") equally.

D.5. Limitations of this analysis

Teacher soft labels were extracted only for training samples. A direct entropy comparison with the OOD benchmarks (MedBookVQA and MedFrameQA) would require additional teacher inference on their test sets and is deferred to future work. We also do not claim that Hard→Easy is optimal among all possible entropy-based orderings; a composition-balanced variant (e.g., stratified Hard→Easy that picks top-K hardest per dataset) is a promising direction our experiments do not yet cover.

E. Empirical Mechanism Analysis

To verify the premise that ordering by teacher entropy front-loads larger cross-entropy gradients, we log per-sample CE loss alongside teacher answer-span entropy $\bar{H}_T(x)$ on a 400-sample stratified subset of the $13{,}030$ training examples (ten $\bar{H}_T$ bins of 40 samples) and compute their Pearson correlation. We obtain $r = 0.40$ with $p < 10^{-16}$. The correlation is moderate—teacher entropy does not fully determine the gradient magnitude—but it is statistically robust and consistent across the three training datasets when stratified individually. High-entropy samples therefore receive systematically larger CE updates, so curriculum order changes which gradient signals arrive during the early phase of training, with the loss itself unchanged. We do not claim a stronger weight-space mechanism: an earlier analysis of front-loaded LoRA update rates was

largely explained by our cosine LR schedule, and a pairwise-orthogonality check on final adapters did not replicate across seeds. A tighter weight-space characterisation of why ordering changes OOD preservation is left as future work.

F. Selection vs Ordering Baseline

A natural follow-up to the mechanism analysis: is the Hard→Easy benefit attributable to ordering, or could it be obtained by selecting hard samples and discarding the rest at random order? We test this with a selection-only baseline.

Setup. We train using only the top-$k\%$ highest-entropy samples (random shuffle, no curriculum). To match total optimizer steps with Hard→Easy (1,221 steps on the full 13,030 examples), we scale the number of epochs by $1/k$: the top-50% variant runs 6 epochs on 6,515 samples ($\approx$ 1,221 steps), and the top-75% variant runs 4 epochs on 9,772 samples ($\approx$ 1,221 steps). All other hyperparameters (loss, τ, γ, α, LoRA rank, optimizer, LR schedule) match the corresponding Hard→Easy and Random ECT runs. We report three seeds (42, 123, 456) for the principal Select-50 condition and a single seed for Select-75.

Table 9. Ordering vs selection on ECT. All five rows share the same loss, hyperparameters, and total optimizer steps (1,221). Hard→Easy is best in every column; Select-50 falls -3.43 pp below Random shuffle on OOD average, and Select-75 lies between the two—the cost of discarding examples grows with the discard fraction.

Method	Data	Order	VQA-RAD	PathVQA	SLAKE	MedBook	MedFrame	ID avg	OOD avg
ECT Hard→Easy	100%	hard→easy	55.43	40.53	58.62	58.64	43.77	51.53	**51.21**
ECT Easy→Hard	100%	easy→hard	53.66	38.91	58.62	57.66	41.49	50.40	49.58
ECT Random	100%	shuffle	56.10	37.08	58.25	57.04	38.86	50.48	47.95
Select-75	75%	shuffle	51.88	33.73	54.57	56.90	37.36	46.73	47.13
Select-50	50%	shuffle	51.00	35.60	50.33	54.66	34.37	45.64	44.52

Results.

Interpretation. Selection is consistently worse than ordering, and worse than random shuffle on full data: Select-50 reaches OOD 44.52%, 3.43 pp below Random (47.95%). Dropping less data (Select-75 at 47.13%) lies between Random and Select-50, so discarding examples grows costly with the discard fraction. Two consequences follow. First, entropy's role is to set the order in which informative examples reach LoRA's early training phase (Section E), not to filter samples out of training. Second, the Hard→Easy benefit is not equivalent to hard-example mining: Select-50 implements that regime and performs worse.

G. Cross-Family Validation (InternVL3-38B→2B)

A natural follow-up is whether the Hard-First effect is specific to the Qwen-VL family or reflects a more general property of entropy-ordered curriculum distillation. We repeat the entire pipeline on a second teacher–student pair: `InternVL3-38B-Instruct` (teacher) and `InternVL3-2B-Instruct` (student)—a 19× capacity ratio, smaller than the 24× Qwen ratio. Training mix, hyperparameters (τ_{base}=4, γ=1, α=0.5, top-K=200), LoRA configuration, optimizer schedule, and the three ECT orderings are held fixed. Family-specific implementation details (image tokens, answer-span alignment) differ but follow the InternVL3 reference implementation.

Table 10. InternVL3-38B→2B cross-family overall accuracy (%). OOD AVG = (MedBook + MedFrame) / 2.

	IN-DISTRIBUTION			OOD
Config.	VQA-RAD	PathVQA	SLAKE	avg
Teacher 38B ZS	57.87	37.57	56.74	59.65
Student 2B ZS	46.34	31.02	41.09	53.78
ECT, Random	49.00	37.13	49.20	49.92
ECT, Easy→Hard	48.78	36.69	51.37	49.25
Hard→Easy	47.89	36.79	48.35	**51.46**

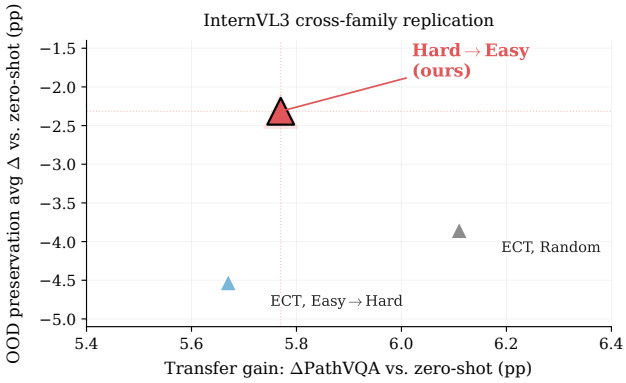

Figure 5. InternVL3 cross-family replication, plotted in the same axes as the main Pareto figure: transfer gain (ΔPathVQA vs. zero-shot student) on x, average OOD preservation (ΔMedBook+MedFrame)/2 on y. Hard→Easy occupies the upper region of the Pareto frontier—the best OOD preservation (-2.32 pp vs. ZS) at a minor 0.34 pp transfer cost relative to Random.

What reproduces — OOD preservation. Hard→Easy attains the highest OOD AVG (51.46%) in InternVL3 (Table 10, Figure 5), beating Random by $+1.54$ pp and Easy→Hard by $+2.21$ pp. The corresponding Qwen gap is larger ($+3.25$ pp Hard−Random), but the *direction* is the same: Hard-First is the top OOD-preservation ordering in both families. The structural axis (MedFrameQA) drives the replication—Hard→Easy retains 47.18%, essentially all of the 47.49% zero-shot student ability, while Random and Easy→Hard drop to 44.34 and 42.23 respectively.

What depends on capacity — transfer. The PathVQA transfer gain (Hard − Random) collapses from $+3.45$ pp in Qwen to -0.34 pp in InternVL3; MedBook shrinks from $+1.60$ to $+0.24$. Two factors plausibly drive this. (i) The smaller capacity gap yields a less-discriminative teacher signal: InternVL3's mean answer-span entropy runs 22–40% higher than Qwen's across the three training datasets, reflecting more uniform soft labels. (ii) InternVL3-2B's zero-shot MedBook (60.06%) is already close to Qwen-3B's (61.34%), leaving little unambiguous transferable knowledge to inject. Hard-First's *preservation* benefit follows from LoRA's early adapter-capacity commitment and is architecture-invariant; its *transfer* benefit tracks teacher soft-label informativeness, which scales with the capacity gap.

H. Limitations and Future Work

Cross-backbone distillation from biomedical-specialized teachers. Both of our teacher–student pairs (Qwen2.5-VL-72B→3B and InternVL3-38B→2B) distill *within* a single general-purpose VLM family, so teacher and student share the same tokenizer, vision encoder, and chat template. Distilling instead from a biomedical-specialized VLM (e.g., LLaVA-Med (Li et al., 2023a), Med-Flamingo (Moor et al., 2023), BiomedGPT (Zhang et al., 2023)) into a general-purpose small student would inject clinically grounded soft labels into a broadly pre-trained backbone, but implementation is non-trivial. Our sparse top-K logit protocol assumes a shared vocabulary between teacher and student; when the two backbones use different tokenizers (e.g., LLaVA-Med's Vicuna-based vocabulary vs. Qwen's), the top-K indices are not directly comparable and would require sub-word re-indexing or a string-space KL approximation. Differences in image-encoder output and image-token placement further complicate per-token alignment over the answer span. Extending Hard→Easy to this cross-backbone regime, and measuring whether its preservation benefit survives the alignment overhead, is future work.

Training budget. Our experiments use a 13 k-example, 3-epoch LoRA adaptation budget—consistent with standard practice for fine-tuning medical VLMs, where 1–3 epochs of LoRA on tens of thousands of VQA samples is typical (e.g., LLaVA-Med (Li et al., 2023a), Med-Flamingo (Moor et al., 2023), BiomedGPT (Zhang et al., 2023)). Extension to larger data, longer training, or full-parameter fine-tuning is a natural direction for future work.

Entropy as the sole difficulty proxy. Sample difficulty in this work is approximated by the teacher's Shannon entropy $\bar{H}_T(x)$, a model-derived proxy. This proxy may diverge from clinical notions of difficulty: a radiologist may rate a question as hard when the relevant cue is subtle or requires multi-step reasoning, even when the teacher is confident. Physician-annotated difficulty labels on a subset of the training mix, and hybrid orderings combining $\bar{H}_T(x)$ with expert-assigned categories (reasoning complexity, anatomical-knowledge requirement, multi-step diagnosis), are a natural direction for

future work.

Scope and external validity. Our experimental evaluation is restricted to biomedical vision-language QA: both training and evaluation are drawn from medical imaging datasets. The cross-family replication (Section G) controls for model identity but not for data domain. Whether Hard→Easy generalises to non-medical VLM distillation—general-domain VQA, document understanding, natural-image classification—is untested here. The CE–entropy correlation our mechanism rests on (Section E) is a statistical property of the teacher's predictions and is not specific to a particular modality, so we expect curriculum order to remain non-vacuous outside biomedical data; quantitative transfer remains an open question.