EM-OPD: EVIDENCE-CONDITIONED MULTI-EXPERT ON-POLICY DISTILLATION FOR PATHOLOGY REPORT GENERATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Multimodal large language models often fail to capture less salient visual evidence required for fine-grained reporting. In contrast, a comprehensive pathology report must integrate heterogeneous evidence from diagnostic patterns, abnormalities, glandular architecture, and cellular morphology. To address this challenge, we propose Evidence-Conditioned Multi-Expert On-Policy Distillation (EM-OPD), a framework that transforms heterogeneous pathological evidence into a unified report generation capability. EM-OPD constructs multiple evidence-conditioned experts equipped with spatial anchors and morphological descriptors. These experts provide complementary perspectives, enabling token-level supervision of student-generated trajectories. By adaptively integrating expert knowledge through balanced joint consolidation, EM-OPD learns to generate pathology reports with improved evidence coverage and diagnostic reliability.

1 INTRODUCTION

Recent multimodal large language models (MLLMs) have achieved strong performance across a broad range of vision-language tasks, spanning visual question answering and open-ended multimodal generation (Singh et al., 2025; Comanici et al., 2025; Qwen Team, 2026). However, globally plausible outputs may fail to condition their reasoning on all task-relevant visual evidence. This issue is especially important for detailed generation, where an output may contain multiple semantic claims supported by different regions and visual patterns. When processing the image as a whole, MLLMs may capture its dominant semantics while overlooking less salient but task-relevant evidence. As a result, their outputs can be broadly correct yet incomplete, imprecise, or partially unsupported. Reliable fine-grained multimodal generation therefore requires models to recognize heterogeneous visual evidence, verbalize the corresponding findings, and integrate them into a coherent output.

Existing training paradigms only partially address this challenge. Supervised fine-tuning (SFT) provides dense token-level supervision through fixed reference responses and can effectively move the model toward the desired output distribution when sufficiently comprehensive training data are available. Such coverage is difficult to obtain for open-ended multimodal generation, where each image is typically paired with only one or a few references despite the large space of valid expressions and possible generation errors. As a result, SFT concentrates supervision on a limited set of reference prefixes and provides no direct corrective signal for many prefixes frequently visited by the student. Reinforcement learning methods such as Group Relative Policy Optimization (GRPO) instead optimize responses sampled from the current policy, thereby targeting the model's actual generation patterns (Shao et al., 2024). However, their outcome-level rewards provide only coarse attribution to the individual claims composing a complex output. This motivates a training paradigm that adaptively supervises student-generated trajectories while retaining dense token-level feedback.

On-policy distillation (OPD) provides such a paradigm: the student samples trajectories from its current policy, while a teacher supplies token-level distribution supervision on the resulting prefixes. Recent OPD methods further improve supervision by equipping the teacher with greater model capacity or privileged information, such as ground-truth answers and localized image crops (Zhao et al., 2026; Yuan et al., 2026). Nevertheless, existing privileged supervision is commonly organized around a single prediction target. This design is effective when the desired output is a concise

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

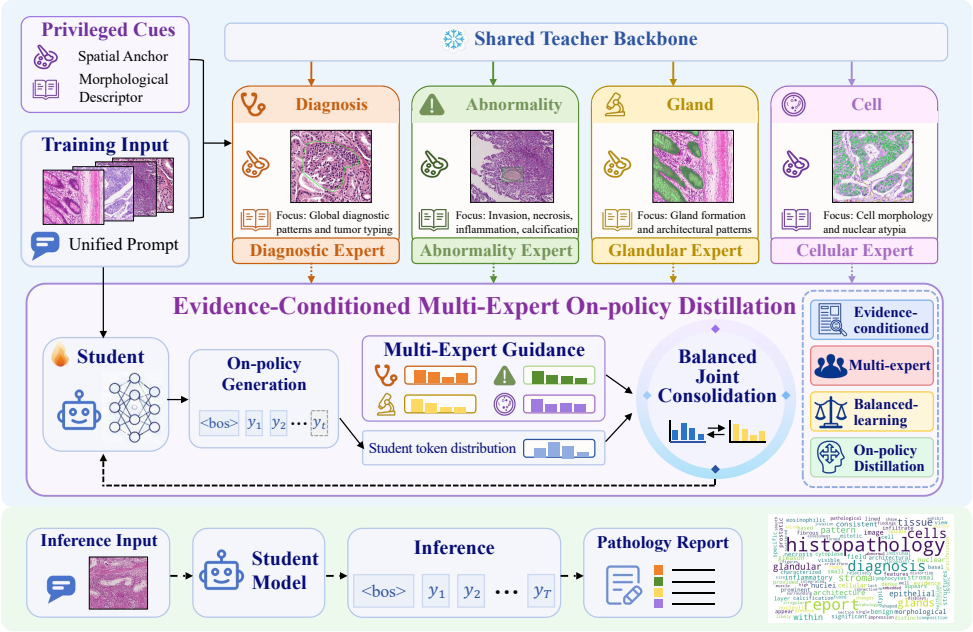


Figure 1: Overview of our framework. During training, multiple evidence-conditioned experts leverage heterogeneous privileged cues to provide complementary token-level supervision, which is adaptively integrated through balanced joint consolidation for on-policy distillation of the student. During inference, the student independently generates pathology reports using only the original image and the unified prompt. The word cloud in the bottom-right qualitatively illustrates the generated reports, highlighting more diverse and fine-grained pathological findings.

answer, but is less suited to open-ended generation in which multiple output components must reflect complementary forms of visual evidence. A single target-conditioned teacher may prioritize the dominant semantics while providing insufficient supervision for less salient evidence, whereas heterogeneous annotations cannot be directly expressed as one unified reference trajectory.

Computational pathology presents a particularly challenging instance of this problem. A pathology report must not only convey an overall diagnosis, but also describe abnormalities, glandular architecture, and distinctive cellular findings. These components depend on different morphological patterns and spatial scales, and their supervision is often distributed across separate annotation sources. Consequently, a model may generate a diagnostically plausible report while omitting or mischaracterizing the fine-grained visual findings needed for a comprehensive and verifiable description. Pathology thus provides a natural setting for studying how heterogeneous privileged evidence can be consolidated into a unified multimodal generation policy.

To this end, we propose **Evidence-Conditioned Multi-Expert On-Policy Distillation** (EM-OPD), a framework that converts heterogeneous privileged pathological evidence into a unified report-generation capability. We instantiate a shared teacher backbone as multiple expert policies, each equipped with task-specific spatial anchors and morphological descriptors. The student generates a free-form report from its current policy while the experts then provide token-level supervision on the student-generated prefixes, transferring complementary diagnostic, abnormality-aware, glandular, and cellular expertise into a single student. At inference time, the student generates reports directly from the original image without privileged evidence or expert policies.

Our contributions are threefold: (1) To the best of our knowledge, we present the first study that applies on-policy distillation to computational pathology. (2) We propose EM-OPD, a novel framework that consolidates heterogeneous privileged pathological evidence into a unified report-generation policy. (3) Extensive experiments demonstrate that EM-OPD generates pathology reports that are both diagnostically accurate and better conditioned on fine-grained pathological evidence.

2 Related Work

2.1 On-Policy Distillation

On-policy distillation trains a student on its self-generated trajectories with token-level teacher supervision, which reduces the train-test mismatch of off-policy distillation (Agarwal et al., 2024; Gu et al., 2024). OPD has since been widely adopted in industrial post-training pipelines, including Qwen3 (Yang et al., 2025), MiMo-V2 (Ma et al., 2026), GLM-5 (Zeng et al., 2026) and DeepSeek-v4 (Xu et al., 2026). Recent work further constructs the teacher and student from different information conditions. On-policy self-distillation (OPSD) conditions the teacher on privileged information, such as verified reasoning traces (Zhao et al., 2026; Penaloza et al., 2026). In multimodal settings, Vision-OPD (Yuan et al., 2026) uses an evidence-centered crop as privileged teacher input and distills regional perception into a full-image policy, while RP-OPSD (Zhu et al., 2026) exploits differences in input resolution as privileged visual supervision.

Concurrent studies have extended OPD to multiple teachers. MOPD (Ma et al., 2026) distills separately optimized domain-specific teachers through prompt-level routing, whereas H-OPD (Yin et al., 2026) combines heterogeneous vision-language and text-only teachers through confidence-aware token-level arbitration. While these methods demonstrate the effectiveness of privileged supervision and multi-teacher distillation, it remains unclear how OPD can integrate heterogeneous expert knowledge corresponding to different visual evidence into a single generation policy.

2.2 Pathological Multimodal Large Language Models

Recent advancements in computational pathology have facilitated the development of MLLMs. Early contributions include PLIP (Huang et al., 2023) and CONCH (Lu et al., 2024a), which leverage CLIP-style contrastive pretraining like CLIP (Radford et al., 2021) and CoCa (Yu et al., 2022). PathoDuet (Hua et al., 2024) further introduces a self-supervised learning framework that jointly exploits multi-scale and multi-stain relationships. Building upon these, PathChat (Lu et al., 2024b) emerges as a comprehensive vision-language generative assistant for pathology. SlideChat (Chen et al., 2024) is introduced as the first MLLM targeting gigapixel slide image analysis. MUSK (Xiang et al., 2025) is another CLIP-based model, but warmed up with unified masked modeling. CPath-Omni (Sun et al., 2025) further advances the field by unifying both patch-level and whole-slide image analysis. TITAN (Ding et al., 2025) is also a slide-level MLLM, achieving state-of-the-art performance for extracting slide representations. Besides, Quilt-LLaVA (Seyfioglu et al., 2024) and PathGen-LLaVA (Sun et al., 2024) are LLaVA-style MLLMs (Liu et al., 2023) based on pathological instruction data. Patho-R1 (Zhang et al., 2026) leverages a combined training scope of supervised finetuning and reinforcement learning to support thoughtful reasoning over pathological tasks. More recently, PathFound (Hua et al., 2026) introduces an agentic multimodal framework that performs evidence-seeking diagnosis through iterative visual inspection and reasoning.

Despite these advances, high-quality pathology reports remain scarce due to the substantial expertise and time of report annotation. Meanwhile, fine-grained pathological evidence may be available through separate annotation sources. Existing pathology MLLMs do not explicitly consolidate such heterogeneous supervision into a unified report-generation policy, which can limit the completeness and evidential conditioning of their outputs.

3 Methods

Pathology report generation requires integrating multiple complementary forms of visual evidence to cover diverse aspects including diagnosis, abnormal findings, glandular architecture, and cellular morphology. Their supervision is distributed across heterogeneous annotation sources and cannot be directly expressed as one unified reference trajectory.

To address this challenge, we propose **Evidence-Conditioned Multi-Expert On-Policy Distillation**. Instead of distilling from a single privileged teacher policy, EM-OPD constructs a *family of evidence-conditioned expert policies* from a shared frozen teacher backbone. Each expert is induced by a distinct type of privileged pathological evidence and provides token-level supervision from a complementary pathological perspective. These expert-specific corrections are applied along trajectories sampled

from the current student policy and are progressively consolidated into a single report-generation model. Our framework consists of three stages, including (1) constructing heterogeneous privileged cues, (2) using these cues to instantiate multiple evidence-conditioned teacher experts, and (3) distilling their complementary knowledge into the student along its on-policy generation trajectories.

3.1 Framework Overview

We consider K training sources that provide complementary privileged evidence for pathology reporting. After preprocessing, the k-th source with N_k images is represented as

$$\mathcal{D}^{(k)} = \left\{ \left(x_i, c_i^{(k)} \right) \right\}_{i=1}^{N_k}, \qquad \mathcal{D} = \bigcup_{k=1}^{K} \mathcal{D}^{(k)}, \tag{1}$$

where x_i denotes the original image, $c_i^{(k)}$ denotes the source-specific privileged cue and $\mathcal{D}$ denotes the aggregated dataset. The internal structure of the cue is detailed in Sec. 3.2.

The **student** model π_θ has access only to the original image x and reporting prompt r, and generates an on-policy trajectory

$$Y^s \sim \pi_\theta(\cdot \mid x, r). \tag{2}$$

At step t, its predictive distribution is

$$p_t = \pi_\theta \left(\cdot \mid x, r, Y^s_{<t} \right). \tag{3}$$

Unlike offline distillation, the **teacher** model does not generate a fixed pseudo-report. Instead, it evaluates the same student-generated prefix with the additional cue

$$q_t^{(k)} = \pi_\phi^{(k)} \left(\cdot \mid x, c^{(k)}, r, Y^s_{<t} \right), \tag{4}$$

where $\pi_\phi^{(k)}$ is the k-th evidence-conditioned expert teacher induced from a shared frozen teacher backbone. Importantly, the privileged cue does not expose the ground-truth report or any sample-specific target response to the teacher.

As illustrated in Fig. 1, their token distributions are aligned to transfer the privileged information into the student. The construction of evidence-conditioned experts and the corresponding on-policy distillation process are detailed in Sec. 3.3, while the consolidation of multiple experts into a unified student is described in Sec. 3.4. Thus, the teacher is privileged during training, whereas the student preserves the standard image-to-report interface without external information. At inference time, the student generates $\hat{Y} \sim \pi_\theta(\cdot \mid x, r)$ using only the original image and reporting prompt. No teacher, privileged cue, source identity, or expert selection is required.

3.2 Privileged Cue Construction

To construct the privileged cues, we first convert the original annotations from each source into a unified two-part format:

$$c^{(k)} = \left(a^{(k)}, d^{(k)} \right), \tag{5}$$

where $a^{(k)}$ is a *spatial anchor* and $d^{(k)}$ is a *morphological descriptor*.

The spatial anchor specifies *where to look*. It identifies the image regions or structures relevant to the corresponding reporting aspect, using visual forms such as bounding boxes, semantic masks, or highlighted regions. This reduces the visual search space and directs the teacher toward localized structures that may otherwise be obscured by the full image. The anchor provides location information without revealing what conclusion should be reported.

The morphological descriptor specifies *what to examine*. It is a textual description of the visual attributes that should be inspected within the relevant regions. While the spatial anchor resolves where the relevant structures are located, the descriptor provides the domain semantics needed to interpret them. It describes the relevant observation dimensions without providing the sample-specific report or target conclusion.

The two components are therefore complementary: the spatial anchor reduces localization ambiguity, whereas the morphological descriptor reduces interpretation ambiguity. Although the original annotation formats differ across sources, they are converted into the same image–anchor–descriptor interface. Table 5 in Appendix A.4 summarizes the construction for each source.

3.3 Evidence-Conditioned On-Policy Distillation

The privileged cues in Sec. 3.2 capture complementary evidence for different components of pathology reporting. Collapsing these cues into a single targeted condition would lead to the student following the dominant semantics while underutilizing the less salient evidence. We therefore formulate the teacher as a set of evidence-conditioned experts, each of which provides token-level supervision from different perspectives of the evidence sources. For evidence source k, the corresponding cue $c^{(k)}$ induces the conditional expert $\pi_\phi^{(k)}$. Its distribution is defined in Eq. 4. All experts share the same architecture and frozen parameters ϕ; their source-specific behavior arises solely from the privileged cue $c^{(k)}$. For a training instance from $\mathcal{D}^{(k)}$, supervision is provided by the corresponding expert $\pi_\phi^{(k)}$.

Given a student rollout Y^s and the token distributions p_t and $q_t^{(k)}$ defined in Sec. 3.1, the selected expert provides token-level supervision on prefixes visited by the evolving student. We align the two distributions using Jensen–Shannon divergence (JSD):

$$m_t^{(k)} = \frac{1}{2}\left(p_t + q_t^{(k)}\right), \tag{6}$$

$$D_{\mathrm{JS}}\left(p_t \| q_t^{(k)}\right) = \frac{1}{2} D_{\mathrm{KL}}\left(p_t \| m_t^{(k)}\right) + \frac{1}{2} D_{\mathrm{KL}}\left(q_t^{(k)} \| m_t^{(k)}\right), \tag{7}$$

where $D_{KL}(\cdot||\cdot)$ denotes the Kullback–Leibler divergence. Accordingly, expert k induces the following student objective:

$$\mathcal{L}_k(\theta) = \mathbb{E}_{\substack{(x,c^{(k)})\sim\mathcal{D}^{(k)}, \\ Y^s\sim\pi_\theta(\cdot|x,r)}} \left[\frac{1}{T} \sum_{t=1}^{T} D_{\mathrm{JS}}\left(p_t \,\middle\|\, \mathrm{sg}[q_t^{(k)}]\right) \right], \tag{8}$$

where $\mathrm{sg}[\cdot]$ denotes stop-gradient and $T = |Y^s|$ denotes the length of the sampled student trajectory. Each evidence source thus defines an expert-specific on-policy distillation objective while sharing the same student parameters θ.

3.4 Balanced Joint Consolidation from Multiple Experts

Balanced joint consolidation. The expert family induces a collection of source-specific objectives $\mathcal{L}(\theta) = (\mathcal{L}_1(\theta), \ldots, \mathcal{L}_K(\theta))$. We jointly consolidate the experts with an equal prior of $1/K$, yielding the balanced joint-consolidation objective of EM-OPD:

$$\mathcal{L}_{\text{EM-OPD}}(\theta) = \frac{1}{K} \sum_{k=1}^{K} \mathcal{L}_k(\theta). \tag{9}$$

This formulation places multi-expert distillation within the multi-objective view of shared-parameter learning (Sener & Koltun, 2018), while making the relative contribution of each expert explicit. In particular, *balanced* denotes equal expert-level weighting, independent of source cardinality or training order.

We realize this objective through stratified source sampling. At every optimization step, we independently draw a mini-batch $\mathcal{B}^{(k)}$ from each source, with $|\mathcal{B}^{(k)}| = B$. Let $\ell_k(\theta;\xi)$ denote the token-averaged on-policy distillation loss for a sampled training instance ξ from source k. The source-wise and consolidated gradient estimators are

$$\widehat{g}_k = \frac{1}{B} \sum_{\xi\in\mathcal{B}^{(k)}} \nabla_\theta \ell_k(\theta;\xi), \qquad \widehat{g}_{\mathrm{bal}} = \frac{1}{K} \sum_{k=1}^{K} \widehat{g}_k. \tag{10}$$

Conditioned on the current student policy, each $\widehat{g}_k$ is an unbiased estimator of the corresponding expert gradient $g_k = \nabla_\theta \mathcal{L}_k(\theta)$. Therefore,

$$\begin{aligned} \mathbb{E}\left[\widehat{g}_{\text{bal}} \mid \theta\right] &= \frac{1}{K}\sum_{k=1}^{K} g_k \\ &= \nabla_\theta \left[\frac{1}{K}\sum_{k=1}^{K} \mathcal{L}_k(\theta)\right] = \nabla_\theta \mathcal{L}_{\text{EM-OPD}}(\theta). \end{aligned} \tag{11}$$

Why not Sequential? An alternative is to distill from experts in separate stages. Consider an update using expert k with learning rate η, the updated student parameters

$$\theta^+ = \theta - \eta g_k. \tag{12}$$

For another expert j, a first-order expansion gives

$$\mathcal{L}_j(\theta^+) = \mathcal{L}_j(\theta) - \eta g_j^\top g_k + \mathcal{O}(\eta^2). \tag{13}$$

When $g_j^\top g_k < 0$, an update that decreases expert k's objective increases expert j's objective to first order. Sequential optimization can therefore overwrite directions favored by earlier experts and introduce order dependence. This **cross-expert interference** is closely related to gradient conflict in shared-parameter multi-objective learning (Chen et al., 2018; Yu et al., 2020; Liu et al., 2021) and motivates keeping all expert objectives active throughout training.

Why not the Original Source Distribution? Balanced consolidation also differs from uniform sampling over the pooled dataset. Let N_k denote the number of training instances associated with source k and define

$$\alpha_k = \frac{N_k}{\sum_j N_j}. \tag{14}$$

Let $\widehat{g}_{\text{pool}}$ denote the stochastic gradient estimator obtained by uniformly sampling training instances from D. Uniform sampling from the pooled dataset then yields the expected gradient

$$\mathbb{E}\left[\widehat{g}_{\text{pool}} \mid \theta\right] = \sum_{k=1}^{K} \alpha_k g_k = \nabla_\theta \left[\sum_{k=1}^{K} \alpha_k \mathcal{L}_k(\theta)\right]. \tag{15}$$

Hence, the original source distribution implicitly determines the optimization weight of each expert: sources with more available annotations contribute more frequently and exert a larger influence on the shared student. This couples optimization importance to annotation availability rather than to the intended expert prior. EM-OPD instead decouples the two by enforcing the explicit $1/K$ expert prior through stratified source sampling, so each evidence source contributes equally at the objective level even when their dataset sizes differ.

Joint optimization keeps all expert objectives active under balanced exposure without assuming conflict-free gradients. Since every $\mathcal{L}_k$ is evaluated on prefixes sampled from the same evolving policy π_θ, the student-induced trajectory distribution also provides a common optimization domain across heterogeneous expert supervision. EM-OPD thus combines *expert-level balancing* across evidence sources with *state-level alignment* through shared on-policy trajectories.

4 Experiments

4.1 Setup

Data. We evaluate our framework on five challenging pathology datasets covering four complementary levels of evidence for pathological report generation: diagnostic, glandular, cellular, and abnormality-level. BreakHis (Spanhol et al., 2015) and GleasonXAI (Mittmann et al., 2025) provide diagnostic supervision: BreakHis is an 8-class breast cancer dataset, while GleasonXAI targets prostate cancer, subdividing Gleason grades 3–5 into 9 fine-grained categories. GlaS (Sirinukunwattana et al., 2017) provides gland-level supervision with gland segmentation and differentiation labels for colon images, and Lizard (Graham et al., 2021) provides cell-level segmentation for colon images across 6 cell types. For abnormality-level supervision, we curate a subset of The Cancer Genome Atlas (TCGA)[1], where pathologists manually annotate the locations of five clinically relevant

[1] https://portal.gdc.cancer.gov

abnormalities: invasion, calcification, inflammation, necrosis, and mitosis. Together, these datasets span heterogeneous visual patterns and annotation formats, enabling comprehensive evaluation of diverse pathological evidence integration. Following Sec. 3.2 and Table 5, we construct spatial anchors and morphological descriptors for all datasets, except that BreakHis lacks spatial anchors due to the absence of cancer region annotations. We further include two out-of-domain datasets in Appendix A.2 to evaluate generalization under domain shifts.

We follow the official split of all datasets for fairness and reproducibility. For the curated TCGA subset, we enforce patient-level separation and check image-level overlap to prevent data leakage.

For all datasets, morphological descriptors are generated by GPT-5.5 (Singh et al., 2025) and manually verified by experienced experts for pathological correctness. Prompts for descriptor generation and teacher models during training are provided in Appendix A.6.

Baselines. We compare our framework with seven representative MLLMs, including general-domain open-source models Qwen3.5-4B, Qwen3.5-9B, and Qwen3.5-27B (Qwen Team, 2026); pathology-specific open-source VLMs PathGen-LLaVA (Sun et al., 2024) and Patho-R1 (Zhang et al., 2026); and state-of-the-art proprietary multimodal models GPT-5.5 and Gemini-3.1-Pro (Comanici et al., 2025), covering different model scales and domain adaptation strategies. Beyond model-level comparisons, we evaluate our OPD strategy against SFT and GRPO. SFT performs supervised fine-tuning on successful generations collected from Gemini-2.5-Pro through multiple sampling attempts, serving as an off-policy distillation baseline, while GRPO serves as a reinforcement learning baseline for assessing policy optimization without our evidence-conditioned distillation mechanism.

We apply our framework under two teacher settings. First, we use the same backbone architecture of the student model as the teacher (OPSD) to evaluate the effectiveness of evidence-conditioned distillation without additional model capacity. Second, we employ a stronger teacher model, Qwen3.5-27B, from the same model family to investigate the benefit of increased teacher capability. For student models, we instantiate our framework with Qwen3.5-4B and Qwen3.5-9B to examine the scalability across different model sizes.

Metrics. Since ground-truth pathological reports are unavailable, conventional NLG metrics such as BLEU and CIDEr cannot be applied. Moreover, a single RoI provides only localized evidence, supporting focused rather than comprehensive conclusions. We therefore use an LLM, Qwen3.5-27B, to extract key insights into structured information and evaluate report coverage and correctness with dataset-specific metrics tailored to each evidence type.

For diagnostic datasets, we report Accuracy (*Acc*) and balanced accuracy (*bAcc*) to address class imbalance. For cellular-level evaluation, which is formulated as a multi-label classification task, we report both micro-F1 and macro-F1. Micro-F1 reflects the overall recognition performance across all cell categories, whereas macro-F1 evaluates performance on each category equally and is therefore more sensitive to minority cell types. For gland and abnormality evaluation, we report Acc and Mention Rate (*MR*). Mention Rate is important because a model may generate generally plausible reports while completely omitting certain pathological aspects.

4.2 Overall Results

Table 1 presents the overall experimental results. As the table shows, existing models tend to capture dominant patterns but may overlook less salient pathological evidence, especially when multiple aspects need to be described simultaneously. Our approach consistently improves performance across diagnostic, glandular, cellular, and abnormality-level evaluations, outperforming both the baseline models and the counterparts trained with SFT or GRPO, demonstrating the effectiveness of consolidating heterogeneous privileged evidence into a unified report generation model.

The improvements are particularly evident at the cell and abnormality levels that require fine-grained visual interpretation, increasing the baseline by 44.97 and 24.59 percentage points respectively. By introducing evidence-conditioned expert supervision, the student model can pay more attention to the subtle but crucial areas. Furthermore, using a stronger teacher backbone provides additional gains over using the same-scale teacher, suggesting that the proposed framework can effectively transfer knowledge from a more capable expert model without introducing additional inference cost. As further shown in Appendix A.2, our method also yields consistent improvements on two

Table 1: Overall results (%) of models. The best performance in each column is **bold** while the second best performance is underlined.

Model	Diagnosis				Gland		Cell		Abnormality	
	BreakHis		GleasonXAI		GlaS		Lizard		TCGA	
	Acc	bAcc	Acc	bAcc	Acc	MR	micro F1	macro F1	Acc	MR
Qwen3.5-4B	17.72	15.42	23.47	22.53	68.75	83.75	36.93	19.62	20.74	36.18
Qwen3.5-9B	29.16	12.06	42.13	20.64	67.50	82.50	40.16	28.75	40.97	70.19
Qwen3.5-27B	31.01	21.75	54.97	29.38	80.00	91.25	45.97	37.64	52.64	79.11
PathGen-LLaVA	22.68	19.50	28.36	15.09	71.25	93.75	63.87	48.58	32.49	69.33
Patho-R1-7B	19.39	15.72	49.88	22.68	67.50	86.25	35.79	29.01	45.39	88.83
Gemini-3.1-Pro	49.80	32.62	45.31	35.69	85.00	**100.00**	79.12	74.87	62.87	89.70
GPT-5.5	55.03	44.22	53.30	25.59	**98.75**	**100.00**	80.00	76.15	56.27	**96.23**
Qwen3.5-4B (SFT)	26.04	13.57	40.01	25.57	85.00	90.00	39.43	32.22	37.34	74.55
Qwen3.5-4B (GRPO)	37.45	16.18	50.89	27.92	85.00	95.00	61.58	52.27	57.51	88.47
Qwen3.5-9B (SFT)	32.06	31.88	55.62	24.55	88.75	93.75	49.03	38.38	42.86	84.05
Qwen3.5-9B (GRPO)	47.25	25.18	57.94	31.84	90.00	95.00	72.76	62.28	61.27	95.72
Qwen3.5-27B (SFT)	48.05	35.71	52.97	35.29	95.00	98.75	69.57	59.03	59.03	86.22
Ours-4B (OPSD)	30.33	14.71	44.50	21.11	85.00	87.50	70.08	62.95	39.38	50.40
Ours-4B (Teacher 27B)	36.37	16.58	48.90	30.95	91.25	98.75	79.33	67.76	43.07	62.29
Ours-9B (OPSD)	51.28	45.44	60.76	33.20	91.25	93.75	80.18	70.92	48.88	74.55
Ours-9B (Teacher 27B)	**57.98**	**52.86**	**61.21**	**46.32**	90.00	97.50	**85.13**	**80.84**	**65.56**	91.01

out-of-distribution datasets, demonstrating its strong generalization capability beyond the training distribution.

Importantly, the distilled student model can even outperform its teacher model on several benchmarks. For example, the 9B student model achieves 57.98% accuracy on BreakHis, outperforming the 27B teacher model's SFT performance of 48.05%. This phenomenon suggests that OPD does not simply force the student to mimic the teacher's generations. Instead, by providing the teacher privileged cues, the student acquires transferable pathological reasoning strategies, including where to focus attention and what morphological characteristics to examine.

4.3 Ablation Study

All ablation studies are conducted under the same setup, with Qwen3.5-9B as the student model and Qwen3.5-27B as the teacher model.

4.3.1 Ablation Study on Privileged Cues

To assess the contribution of different privileged cues, we progressively add the spatial anchors and morphological descriptors to the vanilla OPD framework, with results shown in Fig. 2.

The absence of either component consistently degrades performance, indicating that spatial anchors and morphological descriptors provide complementary supervision. Their relative importance, however, varies with the visual characteristics and reasoning requirements of different pathological evidence sources. For diagnostic-oriented tasks, morphological descriptors contribute more than spatial anchors. Diagnostic regions, such as tumors in histopathological images, often occupy relatively large areas and can be identified from global context, limiting the benefit of additional localization. In contrast, domain-specific morphological knowledge is more important for distinguishing diagnostic patterns and transferring expert pathological knowledge.

For glandular and cellular evidence, spatial anchors are more critical because relevant patterns are distributed across numerous small regions and target structures occupy only limited image areas. Without explicit spatial guidance, the teacher may attend to irrelevant regions or miss structures associated with the target evidence. Spatial anchors therefore reduce visual search ambiguity and improve evidence-conditioned supervision.

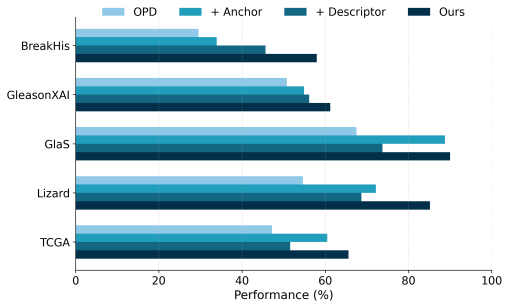


Figure 2: Ablation study on privileged cues. We progressively add the spatial anchors and morphological descriptors to the vanilla OPD framework to evaluate the contribution of each cue and their complementary effects.

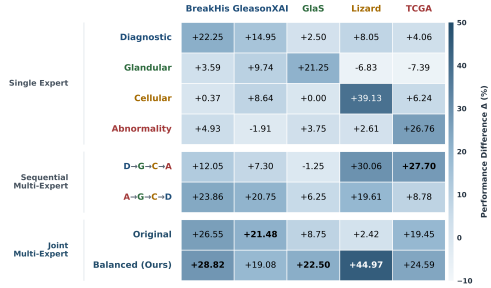


Figure 3: Ablation study on multi-expert supervision. We compare single-expert, sequential multi-expert, and joint multi-expert, including the original source distribution. Values report performance changes relative to Qwen3.5-9B.

4.3.2 Ablation Study on Multi-Expert

To investigate the necessity of consolidating heterogeneous expert knowledge, we compare our balanced joint multi-expert OPD framework with single-expert, sequential multi-expert, and original-distribution variants. The single-expert variants use only one evidence-conditioned expert throughout training, including diagnostic, abnormality, glandular, and cellular experts. The sequential variants introduce multiple experts successively following two orders: Diagnosis→Gland→Cell→Abnormality and Abnormality→Gland→Cell→Diagnosis. The original-distribution variant jointly uses multiple experts while preserving their original source-data proportion. All other settings are kept unchanged. Fig. 3 reports the relative performance changes compared to Qwen3.5-9B.

Single-expert variants perform strongly on their corresponding evidence types, confirming that each expert effectively transfers task-specific privileged knowledge. However, these gains are largely confined to the expert's specialized domain, suggesting that a single privileged perspective cannot capture the heterogeneous evidence required for comprehensive pathology reporting. Sequential multi-expert training also fails to fully consolidate such knowledge. Its performance is highly order-dependent, with the last introduced expert typically receiving the largest gain while knowledge acquired from earlier experts gradually deteriorates. This behavior supports the presence of cross-expert interference.

In contrast, parallel multi-expert training jointly integrates heterogeneous supervision, reducing order dependence and better preserving complementary knowledge across pathological perspectives. Moreover, directly following the original source distribution introduces a clear data-volume bias. The larger GleasonXAI source receives greater improvement, while the relatively smaller glandular and cellular sources are underrepresented. Our balanced strategy mitigates this imbalance and yields more consistent gains across different evidence types. Notably, it also outperforms the diagnostic-only variant in diagnostic accuracy, indicating that complementary pathological supervision not only improves descriptive completeness but also strengthens diagnostic reasoning by enabling richer evidence-conditioned interactions.

5 Conclusions

In this paper, we introduce EM-OPD, an evidence-conditioned multi-expert on-policy distillation framework for pathology report generation. By organizing heterogeneous pathological evidence into multiple evidence-conditioned experts and aligning teacher and student distributions on on-policy trajectories, EM-OPD enables a student model to integrate complementary diagnostic, abnormality, glandular, and cellular knowledge into a unified reporting policy. Future work will explore extending EM-OPD to whole-slide pathology analysis, incorporating richer forms of expert feedback beyond spatial and morphological cues, and investigating more flexible expert conditioning strategies for pathology report generation.

AI USE STATEMENT

In this work, we used generative AI tools for preliminary literature exploration, language polishing, and assistance with the implementation of reinforcement learning components. In particular, generative AI tools were used to assist in adapting and modifying code for our GRPO and OPD implementations based on the open-source GitHub repository verl-project/verl. Besides, we used AI to generate initial morphological descriptors and SFT trajectories, which is described in Sec. 4.1. We did not use generative AI tools for experimental design or data analysis. All AI-assisted content, including text and code, was subsequently manually reviewed and revised by the authors. We take full responsibility for the final content of this work, including all text, claims, analyses, code, and artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

We will open-source the code of our framework to facilitate reproduction. Meanwhile, all the prompts are revealed in the Appendix.

REFERENCES

Rishabh Agarwal, Nino Vieillard, Yongchao Zhou, Piotr Stanczyk, Sabela Ramos Garea, Matthieu Geist, and Olivier Bachem. On-policy distillation of language models: Learning from self-generated mistakes. In *International Conference on Learning Representations*, volume 2024, pp. 21246–21263, 2024.

Andrew A Borkowski, Marilyn M Bui, L Brannon Thomas, Catherine P Wilson, Lauren A DeLand, and Stephen M Mastorides. Lung and colon cancer histopathological image dataset (lc25000). *arXiv preprint arXiv:1912.12142*, 2019.

Nisha Chaudhary, Arpita Rai, Aakash Madhav Rao, Md Imam Faizan, Jeyaseelan Augustine, Akhilanand Chaurasia, Deepika Mishra, Akhilesh Chandra, Varnit Chauhan, and Tanveer Ahmad. High-resolution AI image dataset for diagnosing oral submucous fibrosis and squamous cell carcinoma. *Scientific Data*, 11(1):1050, 2024.

Ying Chen, Guoan Wang, Yuanfeng Ji, Yanjun Li, Jin Ye, Tianbin Li, Ming Hu, Rongshan Yu, Yu Qiao, and Junjun He. Slidechat: A large vision-language assistant for whole-slide pathology image understanding. *arXiv preprint arXiv:2410.11761*, 2024.

Zhao Chen, Vijay Badrinarayanan, Chen-Yu Lee, and Andrew Rabinovich. Gradnorm: Gradient normalization for adaptive loss balancing in deep multitask networks. In *International conference on machine learning*, pp. 794–803. PMLR, 2018.

Gheorghe Comanici, Eric Bieber, Mike Schaekermann, Ice Pasupat, Noveen Sachdeva, Inderjit Dhillon, Marcel Blistein, Ori Ram, Dan Zhang, Evan Rosen, et al. Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. *arXiv preprint arXiv:2507.06261*, 2025.

Tong Ding, Sophia J Wagner, Andrew H Song, Richard J Chen, Ming Y Lu, Andrew Zhang, Anurag J Vaidya, Guillaume Jaume, Muhammad Shaban, Ahrong Kim, et al. A multimodal whole-slide foundation model for pathology. *Nature medicine*, pp. 1–13, 2025.

Simon Graham, Mostafa Jahanifar, Ayesha Azam, Mohammed Nimir, Yee-Wah Tsang, Katherine Dodd, Emily Hero, Harvir Sahota, Atisha Tank, Ksenija Benes, et al. Lizard: A large-scale dataset for colonic nuclear instance segmentation and classification. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 684–693, 2021.

Yuxian Gu, Li Dong, Furu Wei, and Minlie Huang. Minillm: Knowledge distillation of large language models. In *International Conference on Learning Representations*, volume 2024, pp. 32694–32717, 2024.

Shengyi Hua, Fang Yan, Tianle Shen, Lei Ma, and Xiaofan Zhang. PathoDuet: Foundation models for pathological slide analysis of H&E and IHC stains. *Medical Image Analysis*, 97:103289, 2024.

Shengyi Hua, Jianfeng Wu, Tianle Shen, Kangzhe Hu, Zhongzhen Huang, Shujuan Ni, Zhihong Zhang, Yuan Li, Zhe Wang, and Xiaofan Zhang. PathFound: An agentic multimodal model activating evidence-seeking pathological diagnosis. *Medical Image Analysis*, pp. 104200, 2026.

Zhi Huang, Federico Bianchi, Mert Yuksekgonul, Thomas J Montine, and James Zou. A visual–language foundation model for pathology image analysis using medical twitter. *Nature medicine*, 29(9):2307–2316, 2023.

Bo Liu, Xingchao Liu, Xiaojie Jin, Peter Stone, and Qiang Liu. Conflict-averse gradient descent for multi-task learning. *Advances in neural information processing systems*, 34:18878–18890, 2021.

Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. Visual instruction tuning. *Advances in neural information processing systems*, 36:34892–34916, 2023.

Ming Y Lu, Bowen Chen, Drew FK Williamson, Richard J Chen, Ivy Liang, Tong Ding, Guillaume Jaume, Igor Odintsov, Long Phi Le, Georg Gerber, et al. A visual-language foundation model for computational pathology. *Nature Medicine*, 30(3):863–874, 2024a.

Ming Y Lu, Bowen Chen, Drew FK Williamson, Richard J Chen, Melissa Zhao, Aaron K Chow, Kenji Ikemura, Ahrong Kim, Dimitra Pouli, Ankush Patel, et al. A multimodal generative AI copilot for human pathology. *Nature*, 634(8033):466–473, 2024b.

Wenhan Ma, Jianyu Wei, Liang Zhao, Hailin Zhang, Bangjun Xiao, Lei Li, Qibin Yang, Bofei Gao, Yudong Wang, Rang Li, et al. Mopd: Multi-teacher on-policy distillation for capability integration in llm post-training. *arXiv preprint arXiv:2606.30406*, 2026.

Gesa Mittmann, Sara Laiouar-Pedari, Hendrik A Mehrtens, Sarah Haggenmüller, Tabea-Clara Bucher, Tirtha Chanda, Nadine T Gaisa, Mathias Wagner, Gilbert Georg Klamminger, Tilman T Rau, et al. Pathologist-like explainable AI for interpretable Gleason grading in prostate cancer. *Nature communications*, 16(1):8959, 2025.

Emiliano Penaloza, Dheeraj Vattikonda, Nicolas Gontier, Alexandre Lacoste, Laurent Charlin, and Massimo Caccia. Privileged information distillation for language models. *arXiv preprint arXiv:2602.04942*, 2026.

Qwen Team. Qwen3.5: Towards native multimodal agents, February 2026. URL `https://qwen.ai/blog?id=qwen3.5`.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pp. 8748–8763. PmLR, 2021.

Ozan Sener and Vladlen Koltun. Multi-task learning as multi-objective optimization. *Advances in neural information processing systems*, 31, 2018.

Mehmet Saygin Seyfioglu, Wisdom O Ikezogwo, Fatemeh Ghezloo, Ranjay Krishna, and Linda Shapiro. Quilt-llava: Visual instruction tuning by extracting localized narratives from open-source histopathology videos. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13183–13192. IEEE, 2024.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Yang Wu, et al. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024.

Guangming Sheng, Chi Zhang, Zilingfeng Ye, Xibin Wu, Wang Zhang, Ru Zhang, Yanghua Peng, Haibin Lin, and Chuan Wu. Hybridflow: A flexible and efficient rlhf framework. *arXiv preprint arXiv: 2409.19256*, 2024.

Aaditya Singh, Adam Fry, Adam Perelman, Adam Tart, Adi Ganesh, Ahmed El-Kishky, Aidan McLaughlin, Aiden Low, AJ Ostrow, Akhila Ananthram, et al. Openai gpt-5 system card. *arXiv preprint arXiv:2601.03267*, 2025.

Korsuk Sirinukunwattana, Josien PW Pluim, Hao Chen, Xiaojuan Qi, Pheng-Ann Heng, Yun Bo Guo, Li Yang Wang, Bogdan J Matuszewski, Elia Bruni, Urko Sanchez, et al. Gland segmentation in colon histology images: The glas challenge contest. *Medical image analysis*, 35:489–502, 2017.

Fabio A Spanhol, Luiz S Oliveira, Caroline Petitjean, and Laurent Heutte. A dataset for breast cancer histopathological image classification. *Ieee transactions on biomedical engineering*, 63(7): 1455–1462, 2015.

Yuxuan Sun, Yunlong Zhang, Yixuan Si, Chenglu Zhu, Zhongyi Shui, Kai Zhang, Jingxiong Li, Xingheng Lyu, Tao Lin, and Lin Yang. Pathgen-1.6 m: 1.6 million pathology image-text pairs generation through multi-agent collaboration. *arXiv preprint arXiv:2407.00203*, 2024.

Yuxuan Sun, Yixuan Si, Chenglu Zhu, Xuan Gong, Kai Zhang, Pingyi Chen, Ye Zhang, Zhongyi Shui, Tao Lin, and Lin Yang. Cpath-omni: A unified multimodal foundation model for patch and whole slide image analysis in computational pathology. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10360–10371, 2025.

Jinxi Xiang, Xiyue Wang, Xiaoming Zhang, Yinghua Xi, Feyisope Eweje, Yijiang Chen, Yuchen Li, Colin Bergstrom, Matthew Gopaulchan, Ted Kim, et al. A vision–language foundation model for precision oncology. *Nature*, 638(8051):769–778, 2025.

Anyi Xu, Bangcai Lin, Bing Xue, Bingxuan Wang, Bingzheng Xu, Bochao Wu, Bowei Zhang, Chaofan Lin, Chen Dong, Chenchen Ling, et al. Deepseek-v4: Towards highly efficient million-token context intelligence. *arXiv preprint arXiv:2606.19348*, 2026.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, et al. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*, 2025.

Qixiang Yin, Huanjin Yao, Yuchen Cai, Jianghao Chen, Ziyi Wang, Min Yang, Fei Su, and Zhicheng Zhao. H-OPD: Confidence aware heterogeneous multi-teacher multimodal on-policy distillation. *arXiv preprint arXiv:2607.02592*, 2026.

Jiahui Yu, Zirui Wang, Vijay Vasudevan, Legg Yeung, Mojtaba Seyedhosseini, and Yonghui Wu. Coca: Contrastive captioners are image-text foundation models. *arXiv preprint arXiv:2205.01917*, 2022.

Tianhe Yu, Saurabh Kumar, Abhishek Gupta, Sergey Levine, Karol Hausman, and Chelsea Finn. Gradient surgery for multi-task learning. *Advances in neural information processing systems*, 33: 5824–5836, 2020.

Qianhao Yuan, Jie Lou, Xing Yu, Hongyu Lin, Le Sun, Xianpei Han, and Yaojie Lu. Vision-opd: Learning to see fine details for multimodal llms via on-policy self-distillation. *arXiv preprint arXiv:2605.18740*, 2026.

Aohan Zeng, Xin Lv, Zhenyu Hou, Zhengxiao Du, Qinkai Zheng, Bin Chen, Da Yin, Chendi Ge, Chenghua Huang, Chengxing Xie, et al. Glm-5: from vibe coding to agentic engineering. *arXiv preprint arXiv:2602.15763*, 2026.

Wenchuan Zhang, Penghao Zhang, Jingru Guo, Tao Cheng, Jie Chen, Shuwan Zhang, Zhang Zhang, Yuhao Yi, and Hong Bu. Patho-r1: A multimodal reinforcement learning-based pathology expert reasoner. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 28418–28426, 2026.

Siyan Zhao, Zhihui Xie, Mengchen Liu, Jing Huang, Guan Pang, Feiyu Chen, and Aditya Grover. Self-distilled reasoner: On-policy self-distillation for large language models. *arXiv preprint arXiv:2601.18734*, 2026.

Yaowei Zheng, Richong Zhang, Junhao Zhang, Yanhan Ye, and Zheyan Luo. Llamafactory: Unified efficient fine-tuning of 100+ language models. In *Proceedings of the 62nd annual meeting of the association for computational linguistics (volume 3: system demonstrations)*, pp. 400–410, 2024.

Qihui Zhu, Yuchen Wang, Zijian Wen, Tao Zhang, Mengjie Zhang, Yang Liu, Shuangwu Chen, Siying Wu, Jian Yang, and Xiaofeng Jiang. RP-OPSD: Resolution-privileged on-policy self-distillation for multimodal large language models. *arXiv preprint arXiv:2607.24447*, 2026.

A Appendix

A.1 Implementation Details

All models are trained for 3 epochs with $B = 16$ using the AdamW optimizer and a learning rate of 2×10^{-6}. All experiments are conducted on 8 NVIDIA H100 80GB GPUs, and training one pipeline requires approximately 9 hours. For SFT, we leverage the LLaMA-Factory (Zheng et al., 2024) framework. For OPD and GRPO, we leverage the verl (Sheng et al., 2024) framework.

Table 2 summarizes the dataset sizes and the corresponding resampling repetitions used during training.

Table 2: Dataset statistics.

Dataset	Size	Resampling Repetitions
BreakHis	1250	1
GleasonXAI	2442	1
GlaS	85	40
Lizard	168	20
TCGA	1722	2

A.2 Out-of-Domain Evaluation

Table 3: Out-of-domain evaluation on LC25000 and ORCHID.

Model	**LC25000**		**ORCHID**	
	Acc	bAcc	Acc	bAcc
Qwen3.5-4B	42.37	35.66	55.40	43.18
Qwen3.5-9B	70.98	61.47	60.81	47.01
Qwen3.5-27B	78.53	73.33	65.26	46.30
PathGen-LlaVA	52.21	47.99	64.13	45.20
Patho-R1-7B	64.45	61.59	36.56	27.37
Gemini-3.1-Pro	81.90	71.39	67.59	<u>48.71</u>
GPT-5.5	**87.40**	**77.76**	**71.35**	**67.92**
Qwen3.5-9B (SFT)	58.77	46.05	59.61	40.73
Qwen3.5-9B (GRPO)	73.51	73.70	65.02	54.86
Ours-9B	<u>83.79</u>	<u>74.93</u>	<u>69.41</u>	46.67

To further evaluate whether the improvements brought by our method generalize beyond the data distribution used during training, we conduct an additional out-of-domain (OOD) evaluation on the LC25000 dataset (Borkowski et al., 2019) and ORCHID (Chaudhary et al., 2024). LC25000 is a dataset containing images of lung and colon tissues across five diagnostic categories, while ORCHID provides an additional three-class diagnostic benchmark of oral cancer for assessing the generalizability of our method. For both datasets, we exclusively use their test data for inference, with no images or annotations from either dataset involved in training.

As shown in Table 3, our method continues to outperform the vast majority of baselines under this OOD setting, and is surpassed only by the state-of-the-art closed-source model GPT-5.5. These results suggest that the improvement achieved by our method is not merely a consequence of dataset-specific optimization or local adaptation. Instead, teacher-guided learning encourages the student model to identify salient pathological regions and visual evidence that remain informative under distribution shifts. Such transferable pathological understanding enables the model to generate more accurate diagnostic reports on previously unseen data, providing further evidence that the proposed framework improves the model's underlying pathological reasoning capability rather than simply fitting the training distribution.

A.3 Ablation Study on Divergence Computation Strategies

Table 4: Ablation study on divergence computation strategies.

Strategy	Diagnosis				Gland		Cell		Abnormality	
	BreakHis		GleasonXAI		GlaS		Lizard		TCGA	
	Acc	bAcc	Acc	bAcc	Acc	MR	micro F1	macro F1	Acc	MR
ForwardKL	42.95	31.35	**65.44**	33.20	81.25	91.25	74.40	66.05	62.15	87.96
ReverseKL	53.56	27.34	64.51	26.29	76.25	**100.00**	75.59	66.96	59.24	**94.42**
JSD	**57.98**	**52.86**	61.21	**46.32**	**90.00**	97.50	**85.13**	**80.84**	**65.56**	91.01

We further investigate the influence of different divergence computation strategies for aligning the student and teacher distributions. Specifically, we compare Jensen-Shannon divergence with forward KL divergence and reverse KL divergence.

As shown in Table 4, both KL-based alternatives achieve improvements over the student baseline, indicating that token-level distribution alignment itself is beneficial for transferring teacher knowledge. However, Jensen-Shannon divergence consistently provides stronger performance across different pathological evidence sources. This improvement can be attributed to the symmetric property of Jensen-Shannon divergence. In our setting, the teacher and student represent different generation distributions, where neither distribution should be treated as the absolute reference. Forward KL tends to emphasize covering the teacher distribution, while reverse KL encourages the student distribution to concentrate on high-probability teacher modes, which may reduce the diversity of possible pathological descriptions. In contrast, Jensen-Shannon divergence provides a balanced alignment objective by considering both distributions simultaneously.

A.4 Construction of Privileged Cues

Table 5 demonstrates the source-specific construction of the unified image–anchor–descriptor representation.

Table 5: Source-specific construction of the unified image–anchor–descriptor representation.

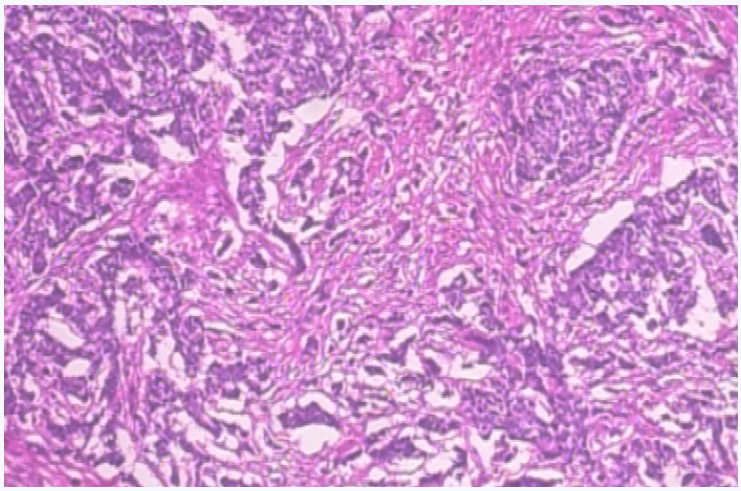
Source	Original annotation	$\mathbf{a}^{(k)}$	$\mathbf{d}^{(k)}$
Diagnosis	Polygon	Contour	Focus on global diagnostic patterns and tumor typing
Abnormality	Polygon	Bounding Box	Focus on invasion, calcification, inflammation, necrosis, and mitosis
Gland	Segmentation Mask	Mask Overlay	Focus on glandular formation and architectural patterns
Cell	Segmentation Mask	Mask Overlay	Focus on cellular morphology and nuclear atypia

A.5 Case Study

We present a case study to demonstrate the effectiveness of the proposed EM-OPD. As shown, the student model initially produces an incorrect diagnosis due to its failure to adequately identify key pathological evidence. Through iterative guidance from multiple experts, the student progressively shifts its attention toward diagnostically relevant cues, including abnormal findings, glandular architecture, and cellular morphology. This refined reasoning process enables the student to interpret comprehensive pathological evidence and ultimately reach the correct diagnosis.

Case Study (The Ground Truth Diagnosis is Ductal Carcinoma)

Given the histopathology image, carefully examine the morphological features and generate a report using cautious pathology language ...



Student(Qwen3.5-9B)'s Original Output:
Based on the H&E-stained histopathology image provided, here is the detailed examination and assessment:

Histopathologic Description

Tissue Architecture:
The image displays a hypercellular proliferation characterized by an intricate network of small, slit-like spaces separated by thin fibrous septa. The overall architecture resembles a s̈taghornv̈ascular pattern or a complex reticular meshwork. There is no evidence of large ducts, lobules, or acini typical of normal breast parenchyma; instead, there is extensive replacement of native stroma by this proliferative process. The structural units appear to be composed of delicate channels lined by flattened endothelium.

Cellular Composition:
The primary constituent appears to be a proliferation of spindle-shaped to plump endothelial cells lining the irregular luminal spaces. These nuclei are elongated but show some variation in shape and orientation within the crowded areas. They lack significant atypia (e.g., prominent nucleoli or coarse chromatin).

Pathological Changes & Features:
* **Vascular Proliferation:** The defining feature is the formation of numerous branching, anastomosing capillaries. Some larger vessels exhibit papillary projections into the lumen (though subtle at this magnification).
* **Nuclear Atypia/Mitoses:** While the cellularity is increased, overt signs of malignancy such as marked nuclear pleomorphism, multinucleation, or frequent mitotic figures are not clearly identifiable in this specific field of view. However, the crowding suggests active growth.

Integrated Impression
The morphological picture demonstrates a well-defined, highly cellular mesenchymal neoplasm featuring a characteristic dual-component structure: alternating zones of hypocellular, hyalinized collagen bundles and hypercellular regions containing branching, dilated vascular channels (c̈ambium layer̈). The presence of leaf-like fronds projecting into cystic spaces—implied by the separation of cellular nodules by compressed stroma—is strongly suggested by the architectural layout.
This constellation of findings—a biphasic tumor with both epithelial-lined (or pseudoepithelial) elements and a florid myxoid/stromal component exhibiting arborizing vasculature—is classic for a Phyllodes spectrum lesion. Given the complexity of the stromal architecture and the prominence of the vascular proliferation without frank sarcomatous transition visible here, it aligns best with the phyllodes tumor.

Most Likely Diagnosis: Phyllodes tumor ✗

Student(Qwen3.5-9B, teacher27B)'s Output after EM-OPD:
Based on the provided histopathology image, here is the morphological report:

Tissue Architecture:
The image displays a breast tissue section characterized by a proliferation of glandular elements embedded within a fibrous stroma. The overall architecture is distorted by a dense, cellular proliferation. The glandular structures are arranged in a complex, cribriform (sieve-like) pattern, where multiple small, round to oval lumens are formed within larger epithelial nests. These structures lack the regular, open, and well-defined acinar organization seen in benign lobules. Instead, they appear as crowded, back-to-back clusters with minimal intervening stroma in some areas. The stroma itself is composed of pink, collagenous connective tissue that appears somewhat compressed and desmoplastic in areas surrounding the epithelial nests.

Cellular Composition:
* **Epithelial Cells:** The predominant cell type consists of epithelial cells forming the glandular structures. These cells exhibit moderate to marked nuclear atypia. The nuclei are enlarged, hyperchromatic (darkly stained), and show variation in size and shape (pleomorphism). The nuclear-to-cytoplasmic ratio is increased. The chromatin pattern appears coarse and clumped. Mitotic figures are not definitively identified in this specific field of view, but the high cellularity suggests active proliferation.
* **Stromal/Connective Tissue:** Between the vascular channels, there is abundant eosinophilic collagenous connective tissue that forms thickened bands separating the clusters of vessels. This creates a distinct biphasic appearance where dense pink fibrous tissue interdigitates with loose, cellular vascular networks.
* **Inflammatory Infiltrate:** Scattered lymphocytes and occasional plasma cells may be seen within the stromal clefts, though they are sparse compared to the dominant neoplastic component.

Abnormal Pathological Changes:
* **Architectural Distortion:** The most prominent abnormality is the loss of normal lobular architecture and the formation of irregular, fused glandular spaces.
* **Cytological Atypia:** The epithelial cells demonstrate clear malignant features, including nuclear enlargement, hyperchromasia, and pleomorphism.
* **Invasion:** The pattern of the glands infiltrating the stroma in a haphazard, non-encapsulated manner is suggestive of an invasive process. The glands are not contained within a distinct ductal or lobular boundary.
* **Calcification/Necrosis/Hemorrhage:** Focal red blood cells are trapped within the slits, indicating recent hemorrhage or congestion common in these lesions. Distinct dystrophic calcifications or geographic necrosis are absent.

Integrated Impression:
The morphological findings are consistent with a malignant epithelial neoplasm of the breast. The presence of irregular, infiltrating glandular structures lined by atypical epithelial cells with high-grade nuclear features strongly supports a diagnosis of invasive carcinoma. The specific architectural pattern of small, angulated glands infiltrating a fibrous stroma is characteristic of **Invasive Ductal Carcinoma (IDC)**.

Final Diagnosis Assessment:
The breast tissue morphology is consistent with a **malignant** lesion.
Most Likely Diagnosis: Ductal Carcinoma ✓

A.6 Prompt and Morphological Descriptor in Use

Here is the prompt for the student model for training and inference.

Student Prompt

```
Given the histopathology image, carefully examine the
    morphological features and generate a report using cautious
    pathology language.

Describe the tissue architecture, including the organization,
    integrity, density, and spatial arrangement of the tissue
    components. When glandular structures are present, characterize
    their morphology, organization, regularity, differentiation,
    and any architectural distortion.
Describe the cellular composition of the image, identifying the
    distinguishable cell types when morphologically supported. For
    each relevant cell population, comment on its appearance,
    relative abundance, distribution, and notable cytological
    features, such as cell size, nuclear morphology, chromatin
    pattern, nuclear-to-cytoplasmic ratio, pleomorphism, and
    mitotic activity.
Assess the abnormal pathological changes, such as calcification,
    necrosis, inflammation, invasion or mitosis.

Describe what is visible, not what you assume. Do not infer
    information beyond the provided field of view, and do not
    overstate uncertain findings. Conclude with a brief integrated
    impression of the main morphological findings and the most
    likely pathological pattern or diagnosis.
```

Here is the morphological descriptor in the BreakHis dataset.

Morphological Descriptor in BreakHis

```
For the final diagnosis, specifically assess whether the breast
    tissue morphology is more consistent with a benign or malignant
    lesion and identify the most likely subtype from the following
    categories:
1.**Adenosis**: Shows an increased number of crowded but
    relatively well-formed acini with recognizable lumina,
    preserved lobular organization, and limited cytological atypia.
2.**Fibroadenoma**: Shows benign ducts that may be round or
    compressed into slit-like spaces within an organized fibrous or
    variably cellular stroma.
3.**Phyllodes tumor**: Shows elongated cleft-like spaces and
    leaf-like stromal projections, usually with greater stromal
    cellularity than fibroadenoma.
4.**Tubular adenoma**: Shows closely packed, relatively uniform
    tubular glands with recognizable lumina, bland epithelial
    cells, and little intervening stroma.
5.**Ductal carcinoma**: Shows atypical cohesive epithelial cells
    forming irregular ducts, nests, cords, tubules, or solid
    sheets, often with infiltrative growth and stromal reaction.
6.**Lobular carcinoma**: Shows small discohesive tumor cells
    infiltrating individually or in single-file arrangements, with
    little gland formation.
7.**Mucinous carcinoma**: Shows clusters, nests, or individual
    tumor cells floating within abundant extracellular mucin.
8.**Papillary carcinoma**: Shows branching papillary epithelial
    structures with fibrovascular cores, often accompanied by
    epithelial crowding, stratification, or atypia.
```

Here is the morphological descriptor in the GleasonXAI dataset.

Morphological Descriptor in GleasonXAI

```
For the final diagnosis, specifically assess whether the prostate
    tissue morphology is more consistent with Gleason pattern 3, 4,
    or 5, and identify the most likely morphological subtype from
    the following categories:
1. **Compressed glands** - Gleason pattern 3: Compressed, angular,
    or irregularly contoured atypical glands remain individually
    recognizable and retain discernible glandular lumina.
2. **Individual glands** - Gleason pattern 3: Separate,
    well-formed malignant glands of variable size infiltrate the
    stroma while maintaining clearly visible lumina.
3. **Cribriform glands** - Gleason pattern 4: A confluent
    epithelial proliferation contains multiple sieve-like glandular
    lumina with no intervening stroma.
4. **Glomeruloid glands** - Gleason pattern 4: A dilated malignant
    gland contains an intraluminal epithelial tuft attached to the
    gland wall at a single point, resembling a renal glomerulus.
5. **Poorly formed glands** - Gleason pattern 4: Small, irregular,
    fused, or poorly defined malignant glands show barely visible,
    slit-like, or absent lumina.
6. **Groups of tumor cells** - Gleason pattern 5: Malignant cells
    form solid nests, clusters, or sheets without recognizable
    glandular differentiation or lumina.
7. **Comedonecrosis** - Gleason pattern 5: Solid, papillary, or
    cribriform tumor structures contain central intraluminal
    necrotic debris, often with karyorrhexis.
8. **Single cells** - Gleason pattern 5: Isolated malignant cells
    infiltrate the stroma individually without forming glands or
    glandular lumina.
9. **Cords** - Gleason pattern 5: Malignant cells infiltrate as
    narrow linear strands or single-cell-thick cords without
    recognizable gland formation.

Notably, pay special attention to the region inside the green
    contour.
```

Here is the morphological descriptor in the GlaS dataset.

Morphological Descriptor in GlaS

```
Specifically assess the glandular structures and determine whether
    their visible morphology is more consistent with benign or
    malignant glands:
**Benign gland**: Typically shows regular and well-defined
    contours, recognizable open lumina, relatively uniform size and
    spacing, orderly epithelial arrangement, preserved polarity,
    and limited cytological atypia.
**Malignant gland**: Typically shows irregular or angulated
    contours, marked variation in size and shape, crowded, fused,
    back-to-back or cribriform architecture, loss of polarity,
    cytological atypia, and possible infiltrative growth or stromal
    destruction.

Use the segmentation map to help locate and distinguish the glands.
```

Here is the morphological descriptor in the Lizard dataset.

Morphological Descriptor in Lizard

```
Specifically identify the following nuclear types that are present:
1. **Neutrophil**: typically has a dark, multilobed or segmented
    nucleus with dense chromatin.
2. **Epithelial**: usually has a relatively large round or oval
    nucleus, often arranged in cohesive sheets or glandular
    structures.
3. **Lymphocyte**: typically has a small, round, dark nucleus with
    very little visible cytoplasm.
4. **Plasma**: usually has an eccentric round nucleus with coarse,
    clock-face-like chromatin.
5. **Eosinophil**: typically has a bilobed nucleus and abundant
    cytoplasmic eosinophilic granules.
6. **Connective tissue**: usually has an elongated,
    spindle-shaped, or flattened nucleus distributed within the
    stromal background.

Use the segmentation map to help locate and distinguish the
    nuclear regions.
```

Here is the morphological descriptor in the TCGA dataset.

Morphological Descriptor in TCGA

```
Specifically assess the following special pathological changes,
    regardless of whether they are present or absent:
1. **Calcification**: Look for basophilic, granular, laminated, or
    amorphous mineral deposits. Describe their location,
    distribution, and relationship to tumor cells, necrotic tissue,
    or stroma.
2. **Necrosis**: Identify areas of cell death, including
    eosinophilic debris, ghost cells, loss of nuclear detail,
    karyorrhexis, or geographic necrosis. Describe whether necrosis
    is focal, extensive, comedo-like, or associated with tumor
    nests.
3. **Inflammation**: Describe inflammatory infiltrates, including
    lymphocytes, plasma cells, neutrophils, histiocytes, or
    granulomatous inflammation. Note whether inflammation is
    peritumoral, intratumoral, stromal, or associated with necrosis.
4. **Invasion**: Evaluate whether cells show infiltrative growth
    into surrounding stroma, vessels, nerves, fat, muscle, or
    adjacent tissue. Describe irregular nests, cords, single-cell
    infiltration, desmoplastic response, or tissue destruction.
5. **Mitosis**: Identify mitotic figures if visible. Describe
    their frequency qualitatively, such as rare, occasional,
    frequent, or brisk. Mention atypical mitotic figures.

Notably, pay special attention to the region inside the green
    bounding box.
```