ReferralSeg: A Multi-Cancer Benchmark for Clinician-Style Language-Guided Segmentation in Histopathology

Roba Al Majzoub[1 [0000−0002−8349−8953]], Ankan Deria[1 [0009−0003−1589−0296]], Numan Saeed[1 [0000−0002−6326−6434]], Salman Khan[1 [0000−0002−9502−1749]], Klaus Hermann Maier-Hein[1,2 [0000−0002−6626−2463]], Shadab Khan[3], and Fahad Shabaz Khan[1 [0000−0002−4263−3143]]

[1] Mohamed Bin Zayed Universityof Artificial Intelligence, Masdar, Abu Dhabi, UAE
[2] German Cancer Research Center (DKFZ), Heidelberg, Germany
[3] ADIA Lab, Abu Dhabi, United Arab Emirates

Abstract. Computational pathology is rapidly moving towards multimodal and conversational systems, yet current methods communicate diagnostic conclusions rather than localized visual evidence supporting them, limiting explainability and evidence-grounded interaction. Pathologists interpret tissue by identifying diagnostically relevant regions through clinically precise descriptions that combine histological features, spatial context, hierarchical assessment, and disambiguation from neighboring structures. This grounded reasoning not only guides diagnosis, but also links clinical decisions to explicit visual evidence. Yet, current methods apply either semantic segmentation or case-level prediction, bypassing this targeted reasoning entirely. We introduce *Referral segmentation* to histopathology: a vision-language task in which models segment a tissue region singled out by clinically grounded descriptions spanning these four reasoning dimensions. We present **ReferralSeg**, a benchmark of image–mask–instruction–quality quadruplets across breast, colon, lung, and prostate cancers, generated through an AI-powered pipeline with cancer-specific prompting schema and two-phase validation, readily extensible to new organs with minimal effort. Evaluation of existing vision-language segmentation models reveals substantial performance gaps, confirming that current methods struggle to follow clinically grounded referral instructions. **ReferralSeg** a foundation for more interactive, explainable and clinically grounded computational pathology systems. The generated dataset will be publicly released.

Keywords: Histopathology · Referral segmentation · Language-grounded segmentation · Dataset · Benchmark

1 Introduction

Computational pathology methods are rapidly moving towards multimodal and conversational systems, to enable more natural interaction between pathologists

and AI. Recent approaches incorporate natural language understanding, text-guided segmentation and report generation to facilitate diagnosis and communicate clinically relevant findings [2, 19, 6, 7]. However, most existing computational pathology methods offer limited support for evidence-grounded interactions. Conventional segmentation models delineate all instances of predefined classes [25, 10], treating every structure equally, while end-to-end approaches directly predict case-level outcomes [8], bypassing region-level analysis. Although recent prompt-based segmentation methods accept points, boxes, or text as input [12], pathology-specific text prompting remains largely restricted to semantic category names or rigid template expressions that cannot uniquely identify one structure within crowded tissue [26, 21, 6]. Consequently, these systems primarily communicate diagnostic conclusions rather than visual evidence supporting them, with predictions that are rarely linked to tissue regions that a pathologist would inspect. Early referral-style approaches were similarly limited to a single cancer type using fixed instruction templates [20].

Recent studies have shown that pathology vision-language models (VLMs) are highly sensitive to textual and visual perturbations, poorly calibrated, and therefore not yet reliable for clinical deployment [13]. More broadly, fluent reasoning traces and strong benchmark performance do not necessarily imply visual grounding [5], highlighting the need for methods that can explicitly connect clinician language to verifiable, localized visual evidence.

Pathologists naturally establish visual grounding by identifying diagnostically informative regions through clinically grounded referral descriptions [24, 15, 16]. Their descriptions draw on four reasoning dimensions at once: what the tissue looks like (histological features such as lumen shape, nuclear polarity, or cytologic atypia), where it sits relative to landmarks (spatial context), how it relates to the overall lesion (hierarchical assessment), and how it differs from similar nearby structures (disambiguation) [9, 1]. These four *reasoning dimensions* uniquely identify a target region and anchor clinical reasoning in verifiable visual evidence. Yet, no existing dataset pairs such descriptions with pixel-level masks.

To address this gap, we introduce *referral segmentation* to computational pathology, a vision-language task in which models segment tissue regions specified through clinically grounded referral descriptions combining the four reasoning dimensions. By explicitly linking natural-language reasoning to pixel-level evidence, referral segmentation provides a foundation for interactive and explainable pathology AI while enabling rigorous evaluation of whether models truly ground their predictions in the visual evidence referenced by clinicians (Fig. 1). Our contributions are fourfold:

- We formalize **referral segmentation** as a task bridging exhaustive segmentation and targeted diagnostic reasoning, grounded in the four reasoning dimensions clinicians use.
- We introduce **ReferralSeg**, a **multi-cancer benchmark** of image-mask-instruction-quality quadruplets spanning breast, colon, lung, and prostate domains (Table 2).

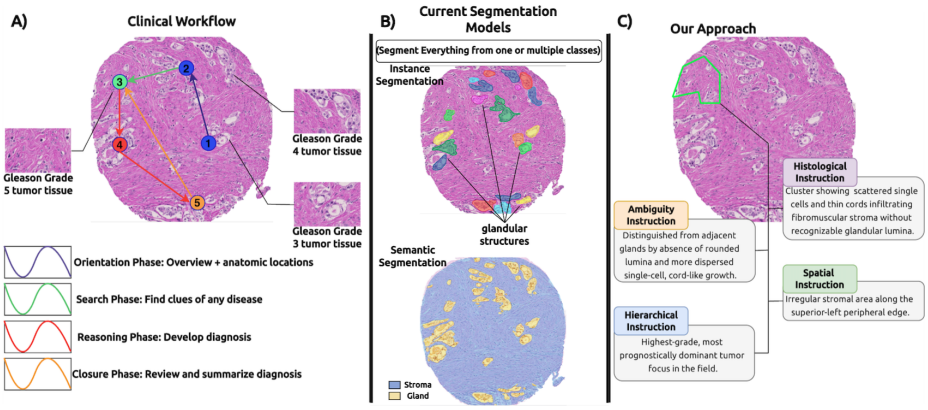

Fig. 1: Comparison between clinician-driven referral reasoning and conventional segmentation paradigms. **A** Pathologists selectively highlight and compare diagnostically relevant structures using reasoning that spans histological, spatial, hierarchical, and disambiguating dimensions. **B** Conventional segmentation exhaustively labels predefined classes without targeting specific regions. **C** Our referral segmentation dataset pairs clinically grounded descriptions with pixel-level masks, aligning computation with diagnostic reasoning.

- We develop a **generalizable AI-powered instruction pipeline** with cancer-specific prompting schema and two-phase validation, releasing the ablation infrastructure as a reproducibility tool.
- We **benchmark existing models** and demonstrate through ablation that combining all four reasoning dimensions is necessary.

2 The ReferralSeg Benchmark

This section describes how we built the benchmark: curating images and masks, generating clinically grounded descriptions, and validating their quality (Fig. 2).

2.1 Dataset Curation and Preprocessing

We collected image-mask pairs from established computational pathology datasets covering gland, tissue, and tumor segmentation as well as cancer grading. The final dataset comprises 2,843 images across four cancer types (Table 1).

Dataset-specific preprocessing was performed to isolate clinically meaningful regions. Multi-class masks (breast and lung) were decomposed into binary regions, individual prostate connected components were treated as separate instances. Representative glands were selected using clinically motivated morphological sampling, including the largest, most elongated, most circular, and a random peripheral gland. [22].

Table 1: Dataset summary. Each cancer type draws from a different source with distinct tissue structures and annotation granularity.

Cancer	Source	Structure	Images	Mag.
Colon	GlaS [18]	Gland inst.	165	20×
Prostate	Arvaniti et al. [4]	Grade regions	641	40×
Breast	TIGER/BCSS [3]	Tissue cls.	1997	20×
Lung	WSSS4LUAD [11]	Tissue cls.	40	40×
Total			**2,843**	—

Table 2: Comparison of BCSS-Ref and our proposed dataset across cancer coverage, resolution, textual diversity, and clinical relevance.

Property	BCSS-Ref	Ours
Resolution	224×224	Multiple resolutions
Cancers included	Breast	Breast, Colon, Lung, Prostate
Textual Variety	✗	✓
Clinical Relevance	✗	✓
Multiple structures	✗	✓

2.2 Instruction Generation

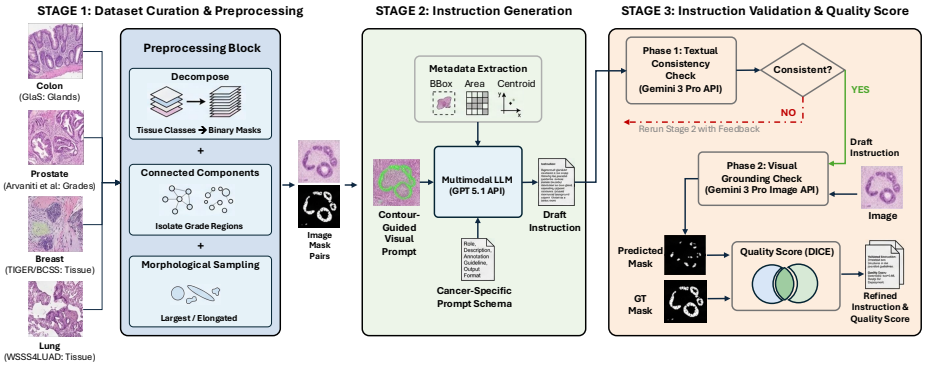

Fig. 2: Overview of the instruction generation pipeline. Please refer to section 2 for a description of the three stages.

For each mask, we extracted spatial metadata (bounding box, centroid, area) and designed cancer-specific prompting schema using terminology from pathology resources [14]. Prompts were built from modular components including clinician role, annotation guidelines, structured disambiguators (e.g., gland orientation or growth patterns), output format, and few-shot examples. This design enables adaptation to new organs by modifying only cancer-specific components.

Instructions were generated using contour-guided visual prompting: H&E images overlaid with high-contrast contours delineating target structures were processed by GPT-5.1 together with cancer-specific prompts and spatial metadata. Every instruction included a core histological description while spatial, hierarchical, and disambiguating cues were added only when needed to uniquely identify the target region (Table 3 for more details.)

2.3 Instruction Validation

Generated instructions were validated in two phases that jointly ensure fidelity of text describing the region, and functional utility. Linguistic correctness alone does not guarantee a description can guide segmentation; conversely, a model might segment correctly from a poor description by exploiting visual shortcuts.

Phase 1: Textual Consistency. Each instruction set was passed with dataset class labels to an independent LLM (Gemini 3 pro preview), which classified the target from text alone. Misclassified instructions were iteratively revised until consistent with ground truth, with 89.19% first-pass accuracy and 100% accuracy after revision.

Phase 2: Visual Grounding. An independent model (Gemini 3 Pro Image Preview) attempted to segment the intended region from the instruction alone. Dice scores guided refinement of cancer-specific prompts through ablation of prompt schema components used to generate the instructions (Fig. 3).

Quality as a Dataset Field. Phase 2 yields a Dice score serving as a proxy for instruction quality: higher scores indicate more discriminative descriptions. **ReferralSeg** is therefore a collection of *image–mask–instruction–quality quadruplets*. This quality score is extensible: future versions can incorporate expert pathologist ratings or inter-annotator agreement through community input on this open dataset (*to be released*).

Ground-truth masks for breast and prostate were repartitioned 7:3 into training and test sets (prostate stratified by tissue class count). Lung masks were similarly divided 7:3. Colon retained its original split. The final ReferralSeg dataset is summarized in Table 3.

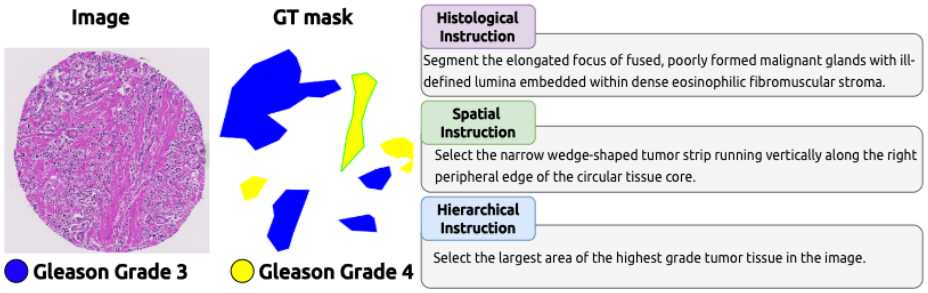

Fig. 3: Example referral instruction for a prostate TMA with multiple tumor grades. The generated text combines histological and hierarchical reasoning to single out the target region with green contour from neighboring regions.

Table 3: Dataset summary. Each cancer type draws from a different source with distinct tissue structures and annotation granularity. Segmentation instructions comprise up to four reasoning components; histological reasoning is present in all cases (100%), while spatial, ambiguity, and hierarchical components appear at varying rates as reported. MIL denotes Mean Instruction Length.

Cancer	**Images**	**Triplets**	**Train/Test**	**MIL (words)**	**%Instruction per component**		
					Spatial	**Ambiguity**	**Hierarchical**
Colon	165	592	297/295	18.7	100	25.17	89.53
Prostate	641	963	667/288	20.6	71.96	54.21	27.1
Breast	1,997	4,064	2,864/1,205	20.6	16.19	46.56	1.65
Lung	40	83	57/26	21.1	37.35	80.72	1.2
Total	**2,843**	**5,702**	**3,885/1,814**	–	–	–	–

Table 4: Ablation: impact of individual reasoning dimensions on segmentation. Combining all dimensions consistently outperforms any single one. See Section 3.1 for details.

Dimension	**Dice**	**IoU**	**Precision**	**Recall**	**Accuracy**
Histological	0.2467	0.1608	0.1751	0.6474	0.5791
Spatial	0.2469	0.1552	0.1744	0.7188	0.5549
Hierarchical	0.2413	0.1527	0.1737	0.6586	0.5756
All Dimensions	**0.2824**	**0.1835**	**0.1965**	**0.7767**	**0.6118**

3 Experiments and Results

We evaluate ReferralSeg from three angles: an ablation testing whether all reasoning dimensions are needed, an overview of quality scores, and a benchmark of existing text-driven segmentation models.

3.1 Ablation: Do All Reasoning Dimensions Matter?

We evaluate the contribution of the three reasoning dimensions present across all images: histological, spatial, and hierarchical. Disambiguating reasoning is conditionally introduced only when the other dimensions cannot uniquely identify the target. Each dimension was tested individually and jointly (Table 4, Fig. 4(A)).

Each dimension achieved a Dice score of 0.24–0.25 individually; combining all raised Dice to 0.28 with gains across all metrics with histological reasoning being the most informative. However, when multiple candidates shared the same histological profile, spatial and hierarchical constraints became essential for disambiguation. This confirms the design premise that singling out one structure among visually similar neighbors requires descriptions drawing on multiple complementary reasoning dimensions.

Table 5: Dice, IoU, gIoU, and cIoU of text-driven segmentation models on our test split.

Model	Colon				Prostate				Breast				Lung			
	Dice	**IoU**	**gIoU**	**cIoU**	**Dice**	**IoU**	**gIoU**	**cIoU**	**Dice**	**IoU**	**gIoU**	**cIoU**	**Dice**	**IoU**	**gIoU**	**cIoU**
MediSee	0.0989	0.0624	-0.1398	-0.0712	0.2976	0.1919	0.1709	0.1806	0.1596	0.0929	0.1377	0.1642	0.1431	0.0800	0.1073	0.1231
BiomedParse	0.2524	0.1595	0.1562	0.0812	0.3553	0.2448	0.5687	0.5457	0.2043	0.1381	0.5957	0.5749	0.1773	0.1182	0.5506	0.5221
Gemini	0.2208	0.1331	0.1623	0.0820	0.5487	0.4537	0.5899	0.5677	0.457	0.33485	0.7147	0.6904	0.4111	0.2861	0.8636	0.8538

3.2 Quality of Instructions

As shown in Fig. 4(C), Quality scores vary considerably across cancer types, with prostate and breast yielding the highest medians and colon the lowest. The poor colon performance reflects the difficulty of localizing tumor-deformed glands, which are often small or peripheral, posing a greater challenge for a general-purpose model without domain-specific training, compared to other cancer types we studied.

3.3 Benchmarking Text-Driven Segmentation Models

We benchmarked three models on the held-out test split (1,842 image–text–mask triplets): MediSee [21], a medical visual reasoning and segmentation model; BiomedParse [26], a biomedical foundation model supporting segmentation; and Gemini 3 pro image preview, a general-purpose multimodal model. All were evaluated with publicly available pretrained weights. We excluded PathSegmentor [6] and LTSE [20] (no public code/weights) and DualProtoSeg [23] (requires fixed class names incompatible with free-form descriptions).

We report four metrics. Dice and IoU measure pixel-level overlap. Generalized IoU (gIoU) [17] and Complete IoU (cIoU) [27], adapted from object detection, evaluate the bounding boxes of predicted and ground-truth masks, with gIoU penalizing spatial displacement and cIoU additionally accounting for centroid distance and aspect ratio. Results are summarized in Table 5.

4 Discussion

Existing models fall short. BiomedParse outperformed MediSee across all metrics, reaching a best Dice of 0.3553 on prostate and 0.25 on colon, despite being trained on histopathology data including GlaS. Its key failure mode is segmenting all visible instances regardless of the instruction, treating referral tasks as exhaustive segmentation. MediSee fared worse, with negative gIoU and cIoU on colon, indicating severe spatial misalignment. Gemini, despite lacking pathology-specific training, achieved higher scores on prostate and breast, suggesting that referral segmentation depends heavily on language-guided reasoning beyond domain-specific pretraining.

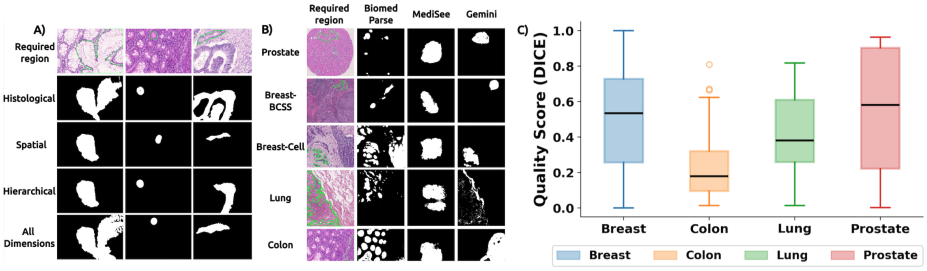

Fig. 4: **A** Ablation showing that individual dimensions often oversegment or mislocalize; the combined instruction constrains the prediction to the correct region. **B** BiomedParse segments all visible structures regardless of the instruction, MediSee produces spatially misaligned predictions, and Gemini better segments the referred region.**C** Distribution of quality scores (Dice) across cancer types. Colon scores are lowest, reflecting the difficulty of describing individual glands in densely packed, morphologically similar tissue.

All four reasoning dimensions are needed. No single dimension provides enough discriminative power (Dice 0.24–0.25); combining them raises performance to 0.28, confirming that referral segmentation requires multiple complementary reasoning dimensions to uniquely identify clinician-referred regions

The pipeline as a generalizable scaffold. The proposed pipeline's modular architecture, including composable prompt fields, means extending the **ReferralSeg** to a new organ requires editing only disease-specific fields, while remaining infrastructure (spatial metadata extraction, contour-guided prompting, two-phase validation) carries over unchanged, lowering the barrier for creating clinically grounded instruction datasets for other modalities.

Implications. These findings reinforce the motivation outlined in the introduction: Referral segmentation exposes the missing capability in current pathology VLMs to ground clinician language in localized visual evidence. Current models default to exhaustive segmentation when given referral instructions, highlighting the need for benchmarks based on clinically grounded, multi-constraint referral descriptions to drive the development of truly evidence-grounded vision-language pathology systems.

5 Limitations and Conclusion

Instruction quality is assessed using automated AI-based evaluation rather than expert pathologist review. The quality score is recorded as a dataset field, however, future versions of this open dataset can incorporate expert pathologist ratings, inter-observer agreement, and whole-slide image evaluation. Although the current pipeline relies on proprietary MLLMs, we mitigate this by releasing prompting schema, metadata extraction code, and ablation infrastructure. By introducing the referral segmentation to computational pathology, this work

bridges natural-language clinical reasoning and pixel-level visual evidence. **ReferralSeg** provides a foundation for developing and evaluating interactive, explainable, and evidence-grounded pathology AI systems that localize the tissue regions underlying their predictions, rather than communicating diagnostic conclusions alone.

References

1. Agosti, V., Munari, E.: Histopathological evaluation and grading for prostate cancer: current issues and crucial aspects. Asian Journal of Andrology **26**(6), 575–581 (2024)
2. Ahmed, F., Sellergren, A., Yang, L., Xu, S., Babenko, B., Ward, A., Olson, N., Mohtashamian, A., Matias, Y., Corrado, G.S., *et al.*: Pathalign: A vision-language model for whole slide images in histopathology. arXiv preprint arXiv:2406.19578 (2024)
3. Amgad, M., Elfandy, H., Hussein, H., Atteya, L.A., Elsebaie, M.A., Abo Elnasr, L.S., Sakr, R.A., Salem, H.S., Ismail, A.F., Saad, A.M., *et al.*: Structured crowdsourcing enables convolutional segmentation of histology images. Bioinformatics **35**(18), 3461–3467 (2019)
4. Arvaniti, E., Fricker, K.S., Moret, M., Rupp, N., Hermanns, T., Fankhauser, C., Wey, N., Wild, P.J., Rueschoff, J.H., Claassen, M.: Automated Gleason grading of prostate cancer tissue microarrays via deep learning. Scientific reports **8**(1), 12054 (2018)
5. Asadi, M., O'Sullivan, J.W., Cao, F., Nedaee, T., Rajabalifardi, K., Li, F.-F., Adeli, E., Ashley, E.: Mirage: The illusion of visual understanding. arXiv preprint arXiv:2603.21687 (2026)
6. Chen, Z., Hou, J., Lin, L., Wang, Y., Bie, Y., Wang, X., Zhou, Y., Chan, R.C.K., Chen, H.: Segment Anything in Pathology Images with Natural Language. arXiv preprint arXiv:2506.20988 (2025)
7. Dai, D., Zhang, Y., Xu, L., Yang, Q., Shen, X., Xia, S., Wang, G.: Pa-llava: A large language-vision assistant for human pathology image understanding. In: 2024 IEEE international conference on bioinformatics and biomedicine (BIBM), pp. 3138–3143 (2024)
8. El Nahhas, O.S., van Treeck, M., Wölflein, G., Unger, M., Ligero, M., Lenz, T., Wagner, S.J., Hewitt, K.J., Khader, F., Foersch, S., *et al.*: From whole-slide image to biomarker prediction: end-to-end weakly supervised deep learning in computational pathology. Nature protocols **20**(1), 293–316 (2025)
9. Feakins, R., Borralho Nunes, P., Driessen, A., Gordon, I.O., Zidar, N., Baldin, P., Christensen, B., Danese, S., Herlihy, N., Iacucci, M., *et al.*: Definitions of histological abnormalities in inflammatory bowel disease: an ECCO position paper. Journal of Crohn's and Colitis **18**(2), 175–191 (2024)
10. Graham, S., Vu, Q.D., Jahanifar, M., Raza, S.E.A., Minhas, F., Snead, D., Rajpoot, N.: One model is all you need: Multi-task learning enables simultaneous histology image segmentation and classification. Medical Image Analysis **83**, 102685 (2023). https://doi.org/10.1016/j.media.2022.102685. https://www.sciencedirect.com/science/article/pii/S1361841522003139

11. Han, C., Pan, X., Yan, L., Lin, H., Li, B., Yao, S., Lv, S., Shi, Z., Mai, J., Lin, J., Zhao, B., Xu, Z., Wang, Z., Wang, Y., Zhang, Y., Wang, H., Zhu, C., Lin, C., Mao, L., Wu, M., Duan, L., Zhu, J., Hu, D., Fang, Z., Chen, Y., Zhang, Y., Li, Y., Zou, Y., Yu, Y., Li, X., Li, H., Cui, Y., Han, G., Xu, Y., Xu, J., Yang, H., Li, C., Liu, Z., Lu, C., Chen, X., Liang, C., Zhang, Q., Liu, Z.: WSSS4LUAD: Grand Challenge on Weakly-supervised Tissue Semantic Segmentation for Lung Adenocarcinoma. In: arXiv (2022)
12. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications **15**(1), 654 (2024). https://doi.org/10.1038/s41467-024-44824-z
13. Majzoub, R.A., Malik, H., Naseer, M., Zaheer, Z., Mahmood, T., Khan, S., Khan, F.: How good is my histopathology vision-language foundation model? a holistic benchmark. arXiv preprint arXiv:2503.12990 (2025)
14. PathologyOutlines.com, Colon adenocarcinoma, https://www.pathologyoutlines.com/topic/colontumoradenocarcinoma.html (2026). Accessed: 2026-02-19.
15. Plass, M., Kargl, M., Nitsche, P., Jungwirth, E., Holzinger, A., Müller, H.: Understanding and Explaining Diagnostic Paths: Toward Augmented Decision Making. IEEE Computer Graphics and Applications **42**(6), 47–57 (2022). https://doi.org/10.1109/MCG.2022.3197957
16. Pohn, B., Kargl, M., Reihs, R., Holzinger, A., Zatloukal, K., Müller, H.: Towards a deeper understanding of how a pathologist makes a diagnosis: Visualization of the diagnostic process in histopathology. In: 2019 IEEE symposium on computers and communications (ISCC), pp. 1081–1086 (2019)
17. Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., Savarese, S.: Generalized intersection over union: A metric and a loss for bounding box regression. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 658–666 (2019)
18. Sirinukunwattana, K., Pluim, J.P., Chen, H., Qi, X., Heng, P.-A., Guo, Y.B., Wang, L.Y., Matuszewski, B.J., Bruni, E., Sanchez, U., *et al.*: Gland segmentation in colon histology images: The glas challenge contest. Medical image analysis **35**, 489–502 (2017)
19. Sun, Y., Zhang, Y., Si, Y., Zhu, C., Zhang, K., Shui, Z., Li, J., Gong, X., Lyu, X., Lin, T., *et al.*: Pathgen-1.6 m: 1.6 million pathology image-text pairs generation through multi-agent collaboration. In: International Conference on Learning Representations, pp. 94611–94653 (2025)
20. Tang, J., Qian, B., Wan, P., Shao, W., Zhang, D.: LTSE: Language-guided Tissue Referring Segmentation in Pathology Images with Adaptive Expert Mixture. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2025, pp. 405–415. Springer Nature Switzerland (2025)
21. Tong, Q., Lu, Z., Liu, J., Zheng, Y., Lu, Z.-M.: Medisee: Reasoning-based pixel-level perception in medical images. In: Proceedings of the 33rd ACM International Conference on Multimedia, pp. 2742–2751 (2025)
22. Trindade, A., Raphael, K., Inamdar, S., Stewart, M., Berkowitz, J., Vegesna, A., Mckinley, M., Benias, P., Kahn, A., Leggett, C., Lee, C.W., Sejpal, D., Rishi, A.: Volumetric laser endomicroscopy features of dysplasia at the gastric cardia in Barrett's oesophagus: results from an observational cohort study. BMJ Open Gastroenterology **6** (2019). https://doi.org/10.1136/bmjgast-2019-000340
23. Vu, A.M., Le, K.P., Vo, T.T., Thach, H., Nguyen, H.H., Yang, D., Huynh, H.H., Nguyen, Q., Pham, T.M., Le, T.-A., *et al.*: DualProtoSeg: Simple and Efficient

Design with Text-and Image-Guided Prototype Learning for Weakly Supervised Histopathology Image Segmentation. arXiv preprint arXiv:2512.10314 (2025)
24. Wang, S., Wu, R., Herndon, C., Liu, Y., Koga, S., Shen, J., Huang, Z.: Pathology-cot: Learning visual chain-of-thought agent from expert whole slide image diagnosis behavior. arXiv preprint arXiv:2510.04587 (2025)
25. Xu, Q., Duan, W., Chen, Z.: Co-Seg: Mutual Prompt-Guided Collaborative Learning for Tissue and Nuclei Segmentation. In: proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2025. Springer Nature Switzerland (2025)
26. Zhao, T., Gu, Y., Yang, J., Usuyama, N., Lee, H.H., Naumann, T., Gao, J., Crabtree, A., Abel, J., Moung-Wen, C., *et al.*: Biomedparse: a biomedical foundation model for image parsing of everything everywhere all at once. arXiv preprint arXiv:2405.12971 (2024)
27. Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., Ren, D.: Distance-IoU loss: Faster and better learning for bounding box regression. In: Proceedings of the AAAI conference on artificial intelligence, pp. 12993–13000 (2020)