

Letter to the Editor

Beyond Visual Consensus: Tiered Reference Framework for AI Cystoscopy Studies

Ahmet Murat Bayraktar, MD; Bilgi İşler, MD

Department of Urology, Konya City Hospital, Konya, Turkey

Corresponding Author:

Ahmet Murat Bayraktar, MD
Department of Urology
Konya City Hospital
Akabe Mah, Adana Çevre Yolu Cad. No.135 Karatay
Konya 42020
Turkey
Phone: 90 332310000 ext 21022
Email: drahmetbayraktar@gmail.com

Related Articles:

Comment on: <https://www.jmir.org/2026/1/e87193>

Comment in: <https://www.jmir.org/2026/1/e103335>

J Med Internet Res 2026;28:e101910; doi: [10.2196/101910](https://doi.org/10.2196/101910)

Keywords: multimodal; large language model; AI; cystoscopy; diagnostic reasoning; finding description; biopsy indication; bladder tumor; artificial intelligence

We read with great interest the study by Shih et al [1], a valuable contribution to the emerging field of artificial intelligence (AI)-assisted cystoscopic diagnosis. Their blinded evaluation of four multimodal large language models across 401 images encompassing 40 cystoscopic finding subcategories provides important insights into current model capabilities. We wish to raise a methodological consideration regarding the reference standard that may inform the interpretation of the reported findings.

The reference standard in this study was established through visual consensus between two urologists, without histopathological confirmation. While interexpert agreement was satisfactory ($\kappa=0.81$), cystoscopic impression alone has well-documented limitations. Cina et al [2] demonstrated that experienced urologists could not reliably distinguish between low- and high-grade papillary lesions endoscopically, with complete grade-stage concordance with histopathology in only 70.3% of cases and a specificity of just 57% for predicting lamina propria invasion. A visually derived reference standard thus carries inherent diagnostic uncertainty.

This concern is particularly relevant for lesion categories central to the study's 7-class task. Carcinoma in situ (CIS) is notoriously difficult to identify under white light cystoscopy; blue light cystoscopy studies have demonstrated that approximately one-third of CIS lesions are missed by white light alone [3]. Similarly, the frequent misclassification of

papilloma as papillary urothelial carcinoma—acknowledged by the authors as reflecting substantial macroscopic overlap [1]—underscores that definitive classification of these entities requires histological evaluation of architectural and cytological features indistinguishable on endoscopic inspection.

The AI-assisted cystoscopic diagnosis literature has converged on histopathological confirmation as the reference standard. Foundational work such as CystoNet was trained and validated on histologically confirmed lesions [4], and a recent systematic review by Hengky et al [5] restricted inclusion to studies using histopathology as the reference standard. This consensus reflects the clinical reality that the categorical distinctions central to bladder lesion classification—low- versus high-grade carcinoma, CIS versus inflammation, papilloma versus carcinoma—ultimately rest on histological criteria and drive subsequent management.

We acknowledge the logistical challenges of obtaining histopathology for every image in a large, multisource dataset, particularly for benign-appearing or nonresected findings. We, therefore, suggest that future benchmarking studies adopt a tiered reference framework: (1) histopathologically confirmed labels for all lesions undergoing biopsy or resection, encompassing the full malignant spectrum; (2) enhanced cystoscopy correlation (blue light or narrow band imaging) as an intermediate standard, particularly for CIS [3]; and (3) consensus visual labels—explicitly flagged as lower confidence—for benign categories unlikely to undergo

biopsy in routine practice. Stratified performance reporting under such a framework would allow readers to separate genuine algorithmic limitations from ambiguity inherent to the reference standard, providing a more clinically meaningful evaluation.

Acknowledgments

The authors acknowledge the use of generative artificial intelligence (Google Gemini) for language editing and proofreading assistance during the preparation of this manuscript.

Funding

The authors declared no financial support was received for this work.

Conflicts of Interest

None declared.

References

1. Shih YC, Wu CY, Huang SW, Tsai CY. Multimodal large language models for cystoscopic image interpretation and bladder lesion classification: comparative study. *J Med Internet Res*. Jan 28, 2026;28:e87193. [doi: [10.2196/87193](https://doi.org/10.2196/87193)] [Medline: [41605505](https://pubmed.ncbi.nlm.nih.gov/41605505/)]
2. Cina SJ, Epstein JI, Endrizzi JM, Harmon WJ, Seay TM, Schoenberg MP. Correlation of cystoscopic impression with histologic diagnosis of biopsy specimens of the bladder. *Hum Pathol*. Jun 2001;32(6):630-637. [doi: [10.1053/hupa.2001.24999](https://doi.org/10.1053/hupa.2001.24999)] [Medline: [11431718](https://pubmed.ncbi.nlm.nih.gov/11431718/)]
3. Grossman HB, Gomella L, Fradet Y, et al. A phase III, multicenter comparison of hexaminolevulinate fluorescence cystoscopy and white light cystoscopy for the detection of superficial papillary lesions in patients with bladder cancer. *J Urol*. Jul 2007;178(1):62-67. [doi: [10.1016/j.juro.2007.03.034](https://doi.org/10.1016/j.juro.2007.03.034)] [Medline: [17499283](https://pubmed.ncbi.nlm.nih.gov/17499283/)]
4. Shkolyar E, Jia X, Chang TC, et al. Augmented bladder tumor detection using deep learning. *Eur Urol*. Dec 2019;76(6):714-718. [doi: [10.1016/j.eururo.2019.08.032](https://doi.org/10.1016/j.eururo.2019.08.032)] [Medline: [31537407](https://pubmed.ncbi.nlm.nih.gov/31537407/)]
5. Hengky A, Lionardi SK, Kusumajaya C. Can artificial intelligence aid the urologists in detecting bladder cancer? *Indian J Urol*. 2024;40(4):221-228. [doi: [10.4103/iju.iju_39_24](https://doi.org/10.4103/iju.iju_39_24)] [Medline: [39555437](https://pubmed.ncbi.nlm.nih.gov/39555437/)]

Abbreviations

AI: artificial intelligence

CIS: carcinoma in situ

Edited by Tiffany Leung; This is a non-peer-reviewed article; submitted 20.May.2026; final revised version received 22.May.2026; accepted 04.Jun.2026; published 18.Jun.2026

Please cite as:

Bayraktar AM, İşler B

Beyond Visual Consensus: Tiered Reference Framework for AI Cystoscopy Studies

J Med Internet Res 2026;28:e101910

URL: <https://www.jmir.org/2026/1/e101910>

doi: [10.2196/101910](https://doi.org/10.2196/101910)

© Ahmet Murat Bayraktar, Bilgi İşler. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 18.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.