Journal of the Canadian Association of Gastroenterology, 2025, **8**, S81–S86
https://doi.org/10.1093/jcag/gwae041
Supplement Article

From data to artificial intelligence: evaluating the readiness of gastrointestinal endoscopy datasets

Sami Elamin*[,1], Shreya Johri[1], Pranav Rajpurkar[1], Enrik Geisler[2], Tyler M. Berzin[2]

[1]Department of Biomedical Informatics, Harvard Medical School, Boston, MA 02115, USA,
[2]Center for Advanced Endoscopy, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02115, USA

*Corresponding author: Sami Elamin, Center for Advanced Endoscopy, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA (selamin@bidmc.harvard.edu).

Abstract

The incorporation of artificial intelligence (AI) into gastrointestinal (GI) endoscopy represents a promising advancement in gastroenterology. With over 40 published randomized controlled trials and numerous ongoing clinical trials, gastroenterology leads other medical disciplines in AI research. Computer-aided detection algorithms for identifying colorectal polyps have achieved regulatory approval and are in routine clinical use, while other AI applications for GI endoscopy are in advanced development stages. Near-term opportunities include the potential for computer-aided diagnosis to replace conventional histopathology for diagnosing small colon polyps and increased AI automation in capsule endoscopy. Despite significant development in research settings, the generalizability and robustness of AI models in real clinical practice remain inconsistent. The GI field lags behind other medical disciplines in the breadth of novel AI algorithms, with only 13 out of 882 Food and Drug Administration (FDA)-approved AI models focussed on GI endoscopy as of June 2024. Additionally, existing GI endoscopy image databases are disproportionately focussed on colon polyps, lacking representation of the diversity of other endoscopic findings. High-quality datasets, encompassing a wide range of patient demographics, endoscopic equipment types, and disease states, are crucial for developing effective AI models for GI endoscopy. This article reviews the current state of GI endoscopy datasets, barriers to progress, including dataset size, data diversity, annotation quality, and ethical issues in data collection and usage, and future needs for advancing AI in GI endoscopy.

Keywords: Artificial intelligence; AI; Endoscopy; Datasets; Machine learning; computer vision; Gastroenterology; Data; Algorithm.

Introduction

Incorporating artificial intelligence (AI) into gastrointestinal (GI) endoscopy represents a promising advancement in gastroenterology practice.[1,2] In clinical research on AI, gastroenterology leads all other medical disciplines with nearly over 40 published randomized controlled trials and many ongoing clinical trials.[3] Advancements such as computer-aided detection algorithms for identifying colorectal polyps have already achieved regulatory approval and are being used in routine clinical care, while numerous other AI applications for GI endoscopy are in advanced stages of development. There is also considerable interest in the potential for computer-aided diagnosis to replace conventional histopathology for diagnosing small colon polyps, facilitating "resect and discard" or "diagnose and leave" approaches.[4–6] Additionally, capsule endoscopy could enable even greater levels of AI automation.[7] These AI advancements, to name a few, aim to improve the efficiency, safety, and diagnostic precision of GI endoscopy.

However, despite the development of many AI models in research settings, their generalizability into real clinical practice and their robustness on different data distributions remains inconsistent. Notably, across the wide spectrum of human disease, GI is being outpaced by other medical fields with regard to the breadth of novel AI algorithms being developed. As of June 2024, only 13 of the 882 Food and Drug Administration (FDA) approved AI models are focussed on GI endoscopy.[8] Furthermore, a large majority of available GI endoscopy image databases are disproportionately focussed on colon polyps, with little/no presentation of the diversity of other endoscopic findings.[9] This bias towards colon polyps is also reflected in the proportion of randomized control trials performed and the FDA-approved AI algorithms for GI.[8]

Several studies have highlighted the link between dataset quality and model accuracy.[10] As such, it is essential for datasets to be of high quality and to represent a wide range of patient demographics, endoscopic equipment types, and most importantly across a much broader range of disease states, to ensure the models developed are effective for a wider range of real-world scenarios for GI endoscopy. In this article, we aim to shed light on the current state of datasets in GI endoscopy, barriers to progress, including dataset size, data diversity, annotation quality, and ethical issues in data collection and usage, and future needs.

Current state of datasets in gastrointestinal endoscopy

Current datasets for endoscopy

In recent years, numerous datasets have been curated in the field of GI. These datasets are critical for training AI models, which can help automate the interpretation of endoscopic images and videos, thereby improving the accuracy and

Table 1. Summary of currently available datasets in the field of gastroenterology, grouped by type of classes.

Disease categories	Year	Dataset name	Country
Colon polyps	2012	ETIS-Larib[12]	France
	2015	CVC-ClinicDB[13]	Spain
	2016	ColonoscopicDS[14]	France
	2016	ASU-Mayo[15]	USA
	2017	CVC-EndoSceneStill[16]	Spain
	2017	Kvasir[17]	Norway
	2019	Kvasir-SEG[18]	Norway
	2020	CP-CHILD[19]	China
	2020	PICCOLO[20]	Spain
	2020	EDD2020[21]	Multiple
	2020	Hyper-Kvasir[22]	Norway
	2021	Kvasir-Sessile[18]	Norway
	2021	KUMC[23]	Multiple
	2021	LDPolypVideo[24]	China
	2021	SUN[25]	Japan
	2022	SUN-SEG[26]	Japan
	2022	ERS[27]	Poland
	2023	PolypGen[28]	France
	2023	ERCPMP[29]	Iran
	2023	Endo-FM[30]	China
	20xx	POLAR[31]	Multiple
	2023	MedVQA-GI[32]	Norway
Anatomical landmarks	2017	Kvasir[17]	Norway
	2020	Hyper-Kvasir[22]	Norway
	2021	Kvasir-Capsule[33]	Norway
	2022	RI-VCE[34]	USA
Upper GI diseases	2017	Kvasir[17]	Norway
	2020	Hyper-Kvasir[22]	Norway
	2020	IPCL[35]	Taiwan
	2022	ERS[27]	Poland
	2023	GastroVision[36]	Multiple
	2023	GA-IM[37]	China
Small intestine diseases	2021	CrohnIPI[38]	France
	2021	Kvasir-Capsule[33]	Norway
	2021	Kvasir-Capsule-SEG[39]	Norway
	2023	AICE[40]	Japan
Large intestine diseases	2017	Kvasir[17]	Norway
	2020	Hyper-Kvasir[22]	Norway
	2022	ERS [27]	Poland
	2023	GastroVision[36]	Multiple
	2023	MedFMC-Endo[41]	China
Bowel prep quality	2017	Nerthus[42]	Norway

efficiency of diagnostics and treatment planning. AI models have demonstrated high accuracy in detecting abnormalities such as polyps, tumours, and ulcerations.[4,11] Table 1 shows a summary of current datasets, grouped by disease category. We note that a majority of the datasets are focussed on colon polyps.

These datasets are spread across ~15 different countries and 26+ hospitals (Figure 1). However, a vast majority of these datasets are focussed on European populations. While the most prevalent disease represented in these datasets was found to be colon polyps, there are 47 other GI diseases also present, spanning upper and lower GI tracts (Figure 2). Furthermore, there are various images showing therapeutic interventions, anatomical landmarks, normal tissue, and bowel preparation quality levels. We note that there is a notable disparity in the number of samples available per dataset (ranging from a few 100s to 100 000s), as well as in image resolutions and clarity. Furthermore, patient ID information is not available in most datasets.

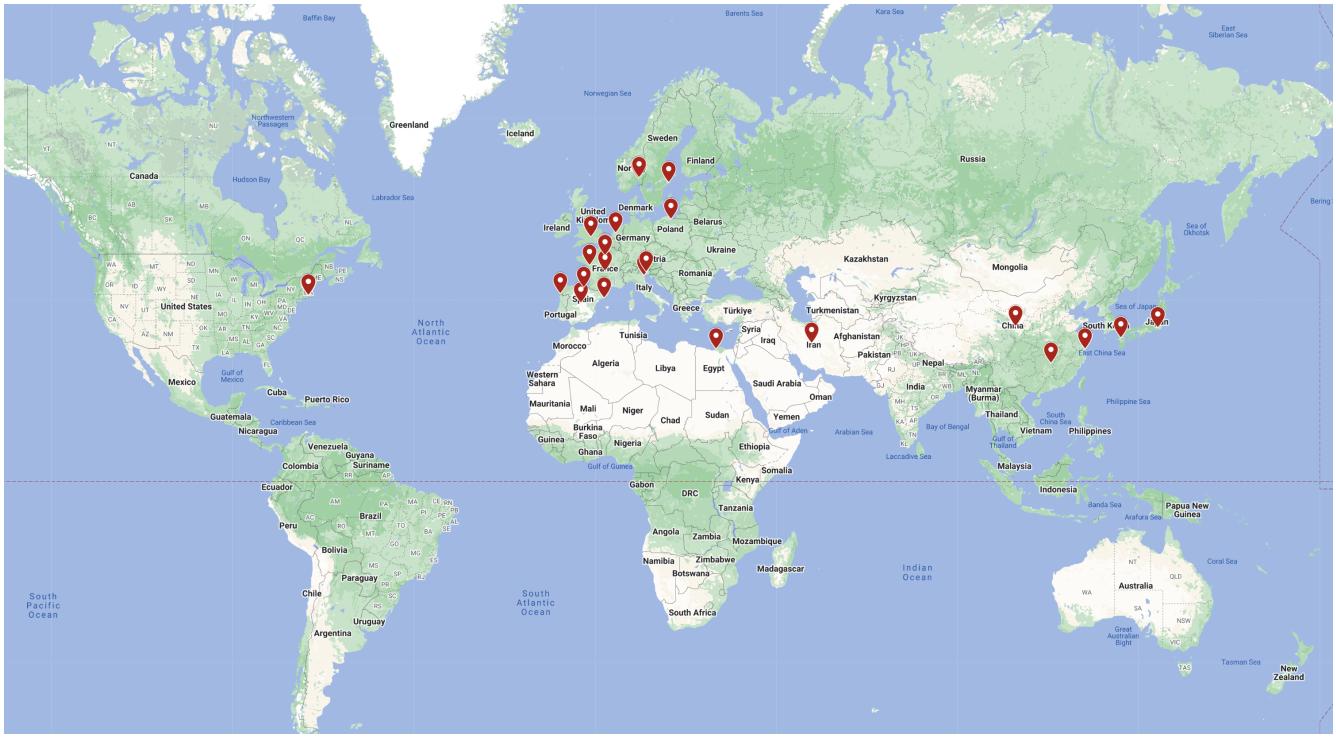

Figure 1. Countries of origin of publicly available GI image and video datasets.

Barriers to progress

Understanding the barriers to progress is crucial for advancing the application of AI in GI endoscopy. By addressing these challenges, we can improve the robustness and generalizability of AI models, ensuring they are effective in diverse clinical settings. Below is a summary table outlining the key barriers to progress.

Collection of diverse and higher-quality datasets

Current datasets exhibit bias towards colon polyps, which limits progress in developing GI models for other diseases. We also note that some datasets have low-quality images, which leads to unreliable AI models. Therefore, larger and high-quality datasets covering a broader set of GI diseases will be needed for progress in training AI models capable of assisting with GI diagnoses in clinical settings.

Furthermore, though data collection efforts have been undertaken across various countries, there are still entire continents that are not well represented (e.g., South America), therefore presenting an opportunity to capture more diversity of patient population in future data collection efforts.

Collection of video datasets

Most of the current endoscopic datasets capture still images only, which limits the development of clinically relevant AI models for diseases such as Barrett's oesophagus and Crohn's disease which require evaluation of more than a single frame to determine disease severity and extent. Video datasets will provide much higher value information to support AI models for assessing these disease states and many others.

Furthermore, most of the few existing video datasets provide only short video snippets that are about ~1–2 mins long and depict the disease. Therefore, the AI models are biased towards assuming that there is a higher likelihood for presence of an abnormality in the image/video provided. This is contrary to clinical settings, where the abnormality is often present in a small part of the entire endoscopic procedure. A collection of full-length endoscopies will be needed to develop better AI models.

Additionally, current endoscopy video datasets typically provide a single label for each short video rather than frame-by-frame labels. This is problematic because, even if most frames in a video snippet depict the disease, it is impossible to evaluate AI models based on their frame-by-frame predictions without such labels. This capability is crucial for real-world deployment. Therefore, future video labelling efforts should focus on collecting frame-by-frame annotations instead of single labels.

Consistency in labelling

Current datasets are labelled in a non-consistent way, leading to difficulty in directly combining datasets for training AI models. For example, some datasets label colon polyp as "polyp," "cancer" or simply "colon-polyp." Other datasets may label a gastric polyp as "polyp." Therefore, when these datasets are combined to train ML models, there is a significant re-labelling effort involved. Therefore, there is a need to adopt a consistent labelling schema, such as Minimum Standard Terminology 3.0,[43] for harmonizing existing datasets and when releasing new ones.

Collection of paired endoscopy reports to automate report generation

Current datasets have been focussed on images or videos, without the accompanying endoscopy report. This limits progress towards the development of AI models for automatic endoscopic report generation. Future efforts towards collecting

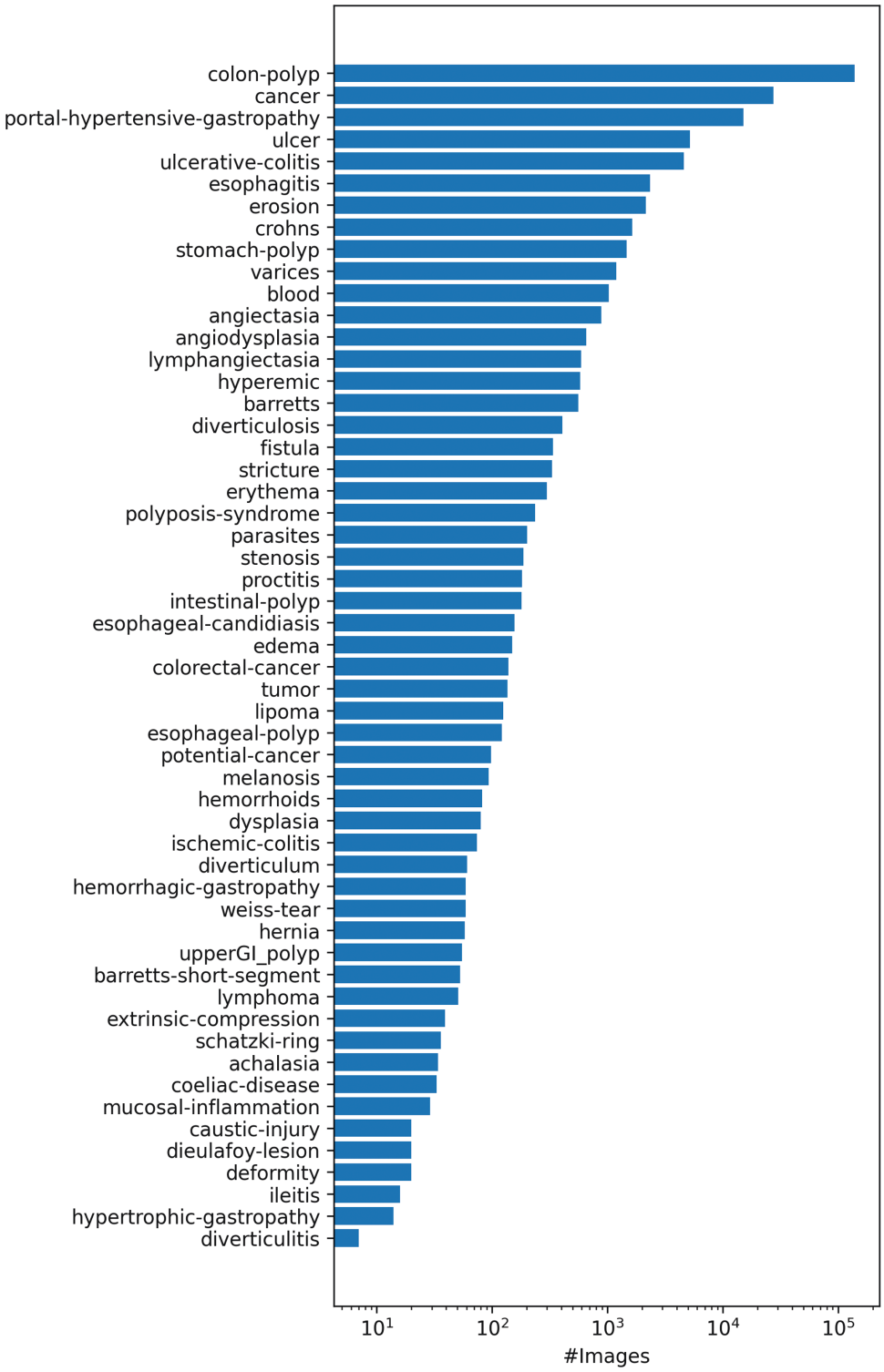

Figure 2. GI diseases covered by datasets.

endoscopy reports will enable development of AI models that can reduce the clinical burden of writing endoscopy reports.

Adhering to ethical standards

Lastly, medical societies and regulatory bodies have taken steps to outline some of the key ethical standards that must be considered during the development and implementation of AI tools in medicine,[44] namely, human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, environmental and societal well-being, and accountability.

Conclusion

We are still in the earliest stages of developing AI tools to support GI endoscopy, and the current shortcomings of available endoscopy datasets are a key barrier to progress and

innovation in this field. Addressing these barriers requires a collaborative effort across the medical and AI communities. By increasing dataset size and diversity, improving annotation quality, adhering to ethical standards, and integrating advanced data types and procedures, we can create more robust and generalizable AI models. Focussing on clinically relevant metrics and leveraging unlabelled data will further enhance the applicability of these models in real-world clinical settings. Through these efforts, we can harness the full potential of AI in GI endoscopy, ultimately improving patient outcomes and streamlining clinical workflows.

Supplementary material

Supplementary material is available at *Journal of the Canadian Association of Gastroenterology* online.

Supplement sponsorship

This article appears as part of the supplement "27th Anniversary Key Topics in Gastroenterology in 2024." The symposium and this supplement were funded by grants from the following sponsors:

- Platinum: Abbvie Canada, Janssen Inc, Pfizer Canada, Takeda Canada, Fresenius-Kabi
- Silver: Boston Scientific, Ferring Pharmaceuticals, Organon, Eli Lilly and Company

Conflicts of interest

T.M.B. is a consultant for Medtronic, Wision AI, Microtech, Magentiq Eye, RSIP Vision, and Boston Scientific. The remaining authors disclose no conflicts. Conflicts of interest disclosure forms (ICMJE) have been collected for all co-authors and can be accessed as supplementary material here

Data availability

There are no new data associated with this article.

References

1. Daniel D. Penrice, Puru Rattan, Douglas A. Simonetto. Artificial intelligence and the future of gastroenterology and hepatology. *Gastro Hep Adv*. 2022;1(4):581–595.
2. Kröner PT, Engels MM, Glicksberg BS, et al. Artificial intelligence in gastroenterology: a state-of-the-art review. *World J Gastroenterol*. 2021;27(40):6794–6824. https://doi.org/10.3748/wjg.v27.i40.6794
3. Han R, Acosta JN, Shakeri Z, Ioannidis JPA, Topol EJ, Rajpurkar P. Randomised controlled trials evaluating artificial intelligence in clinical practice: a scoping review. *Lancet Digit Health*. 2024;6(5):e367–e373. https://doi.org/10.1016/S2589-7500(24)00047-5
4. Ozawa T, Ishihara S, Fujishiro M, Kumagai Y, Shichijo S, Tada T. Automated endoscopic detection and classification of colorectal polyps using convolutional neural networks. *Therap Adv Gastroenterol*. 2020;13(13):1756284820910659. https://doi.org/10.1177/1756284820910659
5. Byrne MF, Chapados N, Soudan F, et al. Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. *Gut*. 2019;68(1):94–100. https://doi.org/10.1136/gutjnl-2017-314547
6. Barua I, Wieszczy P, Kudo SE, et al. Real-time artificial intelligence-based optical diagnosis of neoplastic polyps during colonoscopy. *NEJM Evid*. 2022;1(6):EVIDoa2200003. https://doi.org/10.1056/EVIDoa2200003
7. Sullivan P, Gupta S, Powers PD, Marya NB. Artificial intelligence research and development for application in video capsule endoscopy. *Gastrointest Endosc Clin N Am*. 2021;31(2):387–397. https://doi.org/10.1016/j.giec.2020.12.009
8. U.S. Food and Drug Administration . Accessed August 7 2024. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices
9. Zhu S, Gao J, Liu L, et al. Public imaging datasets of gastrointestinal endoscopy for artificial intelligence: a review. *J Digit Imaging*. 2023;36(6):2578–2601. https://doi.org/10.1007/s10278-023-00844-7
10. He T, Yu S, Wang Z, Li J, Chen Z. From Data Quality to Model Quality: An Exploratory Study on Deep Learning. Asia-Pacific Symposium on Internetware; 2019. Accessed Sep 5, 2024. http://dl.acm.org/citation.cfm?id=3361260
11. Guimarães P, Keller A, Fehlmann T, Lammert F, Casper M. Deep-learning based detection of gastric precancerous conditions. *Gut*. 2020;69(1):4–6. https://doi.org/10.1136/gutjnl-2019-319347
12. Silva J, Histace A, Romain O, Dray X, Granado B. Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer. *Int J Comput Assist Radiol Surg*. 2014;9(2):283–293. https://doi.org/10.1007/s11548-013-0926-3
13. Bernal J, Sánchez FJ, Fernández-Esparrach G, Gil D, Rodríguez C, Vilariño F. WM-DOVA maps for accurate polyp highlighting in colonoscopy: validation vs. saliency maps from physicians. *Comput Med Imaging Graph*. 2015;43(July 2015):99–111. https://doi.org/10.1016/j.compmedimag.2015.02.007
14. Mesejo P, Pizarro D, Abergel A, et al. Computer-aided classification of gastrointestinal lesions in regular colonoscopy. *IEEE Trans Med Imaging*. 2016;35(9):2051–2063. https://doi.org/10.1109/TMI.2016.2547947
15. Tajbakhsh N, Gurudu SR, Liang J. Automated polyp detection in colonoscopy videos using shape and context information. *IEEE Trans Med Imaging*. 2016;35(2):630–644. https://doi.org/10.1109/TMI.2015.2487997
16. Vázquez D, Bernal J, Javier Sánchez F, et al. A benchmark for endoluminal scene segmentation of colonoscopy images. *J Healthc Eng*. 2017;2017:4037190.
17. Konstantin PogorelovSimula Research Laboratory, Kristin Ranheim RandelCancer Registry of Norway, Carsten GriwodzSimula Research Laboratory, Sigrun Losada EskelandVestre Viken Hospital Trust, Thomas de LangeCancer Registry of Norway, Dag JohansenUiT-The Arctic University of Norway, et al. KVASIR. A Multi-Class Image Dataset for Computer Aided Gastrointestinal Disease Detection . *Association for Computing Machinery Digital Library*. Accessed September 2, 2024. https://dl.acm.org/do/10.1145/3193289/
18. Jha, D, Smedsrud PH, Riegler MA, et al. 'Kvasir-Seg: A Segmented Polyp Dataset'. *MultiMedia Modeling: 26th International Conference, MMM 2020, Daejeon, South Korea, January 5–8, 2020, Proceedings, Part II* 26, 2020, pp. 451–462.
19. Wang W, Tian J, Zhang C, Luo Y, Wang X, Li J. An improved deep learning approach and its applications on colonic polyp images detection. *BMC Med Imaging*. 2020;20(1):1–14.
20. Sánchez-Peralta LF, Pagador JB, Picón A, et al. PICCOLO white-light and narrow-band imaging colonoscopic dataset: a performance comparative of models and datasets. *Appl Sci*. 2020;10(23):8501. https://doi.org/10.3390/app10238501
21. Ali S, Braden B, Lamarque D, et al. *Endoscopy Disease Detection and Segmentation (EDD2020)*. IEEE DataPort; 2020. https://doi.org/10.21227/F8XG-WB80
22. Borgli H, Thambawita V, Smedsrud PH, et al. HyperKvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy. *Sci Data*. 2020;7(1):1–14.
23. Li K, Fathan MI, Patel K, et al. Colonoscopy polyp detection and classification: Dataset creation and comparative evaluations.

PLoS One. 2021;16(8):e0255809. https://doi.org/10.1371/journal.pone.0255809
24. Ma Y, Chen X, Cheng K, Li Y, Sun B. *LDPolypVideo Benchmark: A Large-Scale Colonoscopy Video Dataset of Diverse Polyps*. Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. 2021; 387–396.
25. Misawa M, Kudo SE, Mori Y, et al. Development of a computer-aided detection system for colonoscopy and a publicly accessible large colonoscopy video database (with video). *Gastrointest Endosc*. 2021;93(4):960–967.e3. https://doi.org/10.1016/j.gie.2020.07.060
26. Ji G-P, Xiao G, Chou Y-C, et al. Video polyp segmentation: a deep learning perspective. *Mach Intell Res*. 2022;19(6):531–549. https://doi.org/10.1007/s11633-022-1371-y
27. Cychnerski J, Dziubich T, Brzeski A. ERS: a novel comprehensive endoscopy image dataset for machine learning, compliant with the MST 3.0 specification. arXiv [cs.CV]. 2022. Available: http://arxiv.org/abs/2201.08746
28. Ali S, Jha D, Ghatwary N, et al. A multi-centre polyp detection and segmentation dataset for generalisability assessment. *Sci Data*. 2023;10(1):1–17.
29. Forootan M, Rajabnia M, Mafi AR, et al. ERCPMP-v5: An Endoscopic Image and Video Dataset for Recognition of Colorectal Polyps Morphology and Pathology. *Mendeley Data*. 2024;6. https://doi.org/10.17632/7grhw5tv7n.6
30. Wang Z, Liu C, Zhang S, Dou Q. *Foundation Model for Endoscopy Video Analysis via Large-Scale Self-supervised Pre-train*. Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. 2023; 101–111.
31. Dekker E. *ClinicalTrials.gov. Academisch Medisch Centrum - Universiteit van Amsterdam (AMC-UvA) NCT03822390*. Accessed September 2, 2024. https://clinicaltrials.gov/ct2/show/NCT03822390
32. Ionescu B, Müller H, Drăgulinescu A-M, et al. *Overview of the ImageCLEF 2023: Multimedia Retrieval in Medical, Social Media and Internet Applications*. Experimental IR Meets Multilinguality, Multimodality, and Interaction. 2023; 370–396.
33. Smedsrud PH, Thambawita V, Hicks SA, et al. Kvasir-Capsule, a video capsule endoscopy dataset. *Sci Data*. 2021;8(1):1–10.
34. Charoen A, Guo A, Fangsaard P, et al. Rhode Island gastroenterology video capsule endoscopy data set. *Sci Data*. 2022;9(1):1–6.
35. García-Peraza-Herrera LC, Everson M, Lovat L, et al. Intrapapillary capillary loop classification in magnification endoscopy: open dataset and baseline methodology. *Int J Comput Assist Radiol Surg*. 2020;15(4):651–659. https://doi.org/10.1007/s11548-020-02127-w
36. Jha D, Sharma V, Dasu N, et al. GastroVision: A multi-class endoscopy image dataset for computer aided gastrointestinal disease detection. ICML Workshop on Machine Learning for Multimodal Healthcare Data (ML4MHD 2023).
37. Yang J, Ou Y, Chen Z, et al. A benchmark dataset of endoscopic images and novel deep learning method to detect intestinal metaplasia and gastritis atrophy. *IEEE J Biomed Health Inform*. 2023;27(1):7–16. https://doi.org/10.1109/JBHI.2022.3217944
38. de Maissin A, Vallée R, Flamant M, et al. Multi-expert annotation of Crohn's disease images of the small bowel for automatic detection using a convolutional recurrent attention neural network. *Endosc Int Open*. 2021;9(7):E1136–E1144. https://doi.org/10.1055/a-1468-3964
39. Jha D, Tomar N, Ali S, et al. 'Nanonet: Real-Time Polyp Segmentation in Video Capsule Endoscopy and Colonoscopy'. *Proceedings of the IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS)*, IEEE, 2021, pp. 37–43.
40. Yokote A, Umeno J, Kawasaki K, et al. Small bowel capsule endoscopy examination and open access database with artificial intelligence: the SEE-artificial intelligence project. *DEN Open*. 2024;4(1):e258. https://doi.org/10.1002/deo2.258
41. Wang D, Wang X, Wang L, et al. A real-world dataset and benchmark for foundation model adaptation in medical image classification. *Sci Data*. 2023;10(1):1–9.
42. Pogorelov K, Randel KR, de Lange T, et al. *Nerthus: A Bowel Preparation Quality Video Dataset*. Association for Computing Machinery; 2017. https://doi.org/10.1145/3193165
43. Aabakken L, Rembacken B, LeMoine O, et al. Minimal standard terminology for gastrointestinal endoscopy - MST 3.0. *Endoscopy*. 2009;41(8):727–728. https://doi.org/10.1055/s-0029-1214949
44. European Parliament: Directorate-General for Parliamentary Research Services, Lekadir K, Quaglio G, Tselioudis Garmendia A, Gallin C. *Artificial intelligence in healthcare – Applications, risks, and ethical and societal impacts*, European Parliament, 2022. Accessed September 2, 2024. https://data.europa.eu/doi/10.2861/568473