Revista Española de Enfermedades Digestivas
The Spanish Journal of Gastroenterology

Title:
An interpretable artificial intelligence system for measuring the size of small polyps (< 10 mm)

Authors:
Yong Li, Fujun Li, Guanghui Lian, Yanliang Hou, Jiefei Chen, Shizhe Li, Chuquan Cai, Bin Liu, Yu Wu, Xiaowei Liu

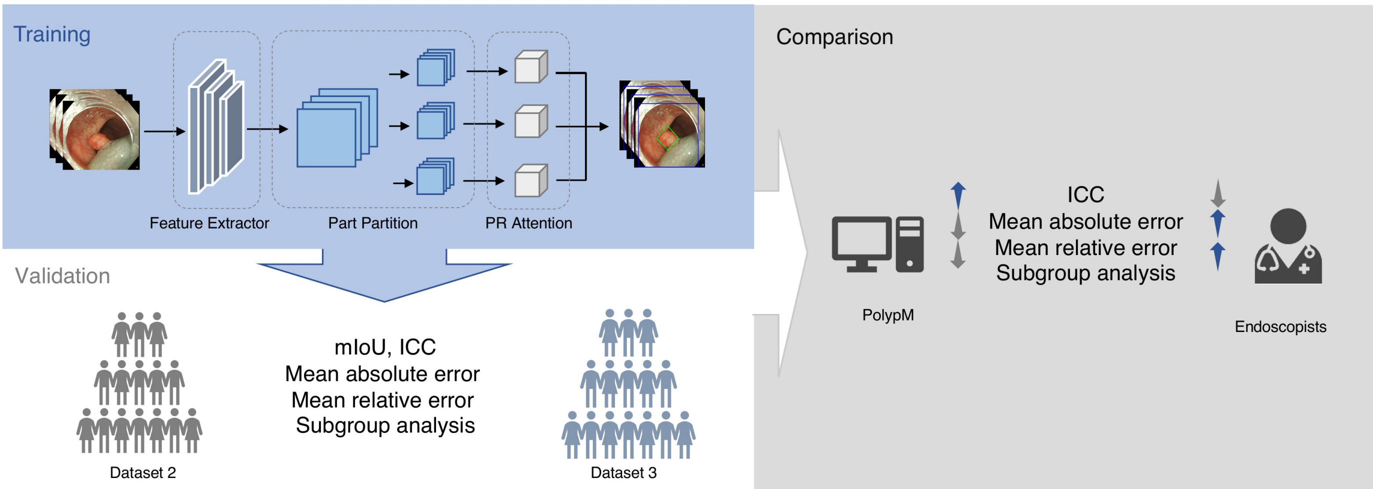
DOI: 10.17235/reed.2025.11539/2025
Link: [PubMed (Epub ahead of print)]

Please cite this article as:
Li Yong, Li Fujun, Lian Guanghui, Hou Yanliang, Chen Jiefei, Li Shizhe, Cai Chuquan, Liu Bin, Wu Yu, Liu Xiaowei. An interpretable artificial intelligence system for measuring the size of small polyps (< 10 mm). Rev Esp Enferm Dig 2025. doi: 10.17235/reed.2025.11539/2025.

Revista Española de Enfermedades Digestivas
The Spanish Journal of Gastroenterology

An interpretable artificial intelligence system for measuring the size of small polyps (<10 mm)

Results presentation



XXX, et al.

Revista Española de Enfermedades Digestivas (REED)
The Spanish Journal of Gastroenterology

Revista Española de Enfermedades Digestivas
The Spanish Journal of Gastroenterology

Revista Española de Enfermedades Digestivas
The Spanish Journal of Gastroenterology

An interpretable artificial intelligence system for measuring the size of small polyps (< 10 mm)

Short running title: A system measuring polyps' size

Yong Li, MD[1,2], Fujun Li, MD, PhD[1,2], Guanghui Lian, MD, PhD[1,2], Yanliang Hou, MD[1,2], Jiefei Chen, MD, PhD[3], Shizhe Li, MD[1,2], Chuquan Cai, MD[4], Bin Liu, MD[5], Yu Wu, MD[1,2*], Xiaowei Liu, MD, PhD[1,2,6]

1. Department of Gastroenterology, Xiangya Hospital, Central South University, Changsha, Hunan 410008, China
2. Hunan International Scientific and Technological Cooperation Base of Artificial Intelligence Computer Aided Diagnosis and Treatment for Digestive Disease, Xiangya Hospital, Central South University, Changsha, Hunan 410008, China
3. Department of Laboratory Medicine, Xiangya Medical School Central South University, Changsha, Hunan 410008, China
4. Department of Gastroenterology, Changsha Eighth Hospital, Changsha, Hunan 410008, China
5. Department of Gastroenterology, Xiangtan Central Hospital, Changsha, Hunan 410008, China
6.Research Center for Geriatric Disorder, Xiangya Hospital, Central South University, Changsha, Hunan 410008, China

*Corresponding author
E-mail addresses: wuyu527@csu.edu.cn (Yu Wu)

Abbreviation： AI, artificial intelligence; CI, confidence interval; PRA-Net, Part Relation Attention Network; IoU, Intersection over Union; IQR, interquartile range; ICC, intraclass correlation coefficient; BBPS, Boston bowel preparation scale; MAE, mean absolute error; MRE, mean relative error

Statements & Declarations

Author contributions

Study concept and design: YL, YW, XWL. Acquisition of data: FJL, GHL, YLH, CQC, BL, XWL. Analysis and interpretation of data: YL, SZL. Statistical analysis: YL, JFC. Drafting of the manuscript: YL, YW. Critical revision of the manuscript for important intellectual content: YW, XWL. All authors approved the final version of the manuscript, including the authorship list.

Declaration of interest

Yong Li, Fujun Li, Guanghui Lian, Yanliang Hou, Jiefei Chen, Shizhe Li, Chuquan Cai, Bin Liu, Yu Wu and Xiaowei Liu have no relevant financial or non-financial interests to disclose.

Artificial intelligence

The authors declare that they did not use artificial intelligence (AI) or any AI-assisted technologies in the elaboration of the article.

Data availability statement

All the data included in this study are available upon request by contacting the corresponding author.

Ethics approval

This study was approved by the Ethical Committee of Xiangya Hospital Central South University (ID:2022033103), Changsha Eighth Hospital (ID:2023022301), Xiangtan Central Hospital (ID:2023031701). The need for written informed consent was obtained from all participants. The study was conducted in accordance with the principles of the Declaration of Helsinki and current guidelines.

Lay summary

This study developed PolypM, an AI system that automates accurate measurement of colon polyp size—especially challenging for small polyps (<10 mm). Using dual models to segment polyps and measurement-assisting transparent caps, PolypM underwent extensive training and validation with thousands of endoscopic images. In multicenter external validation, it consistently outperformed endoscopists, achieving significantly lower sizing errors (both absolute and relative) regardless of polyp shape, pathology, or physician experience. By providing reliable, automated measurements, PolypM addresses critical limitations of manual assessment, directly supporting better clinical decisions about polyp treatment and follow-up timing.

Abstract

Background: Precise polyp size measurement is vital for treatment and follow-up strategy determination, with small polyps (<10 mm) presenting particular challenges. This study aimed to develop an artificial intelligence system, PolypM, for automated polyp size measurement to enhance clinical decision-making.

Methods: PolypM, comprising two models, was designed for automatic segmentation of transparent caps and polyps. It was trained on 6486 endoscopic images (Dataset 1), validated on 675 images (Dataset 2), and compared to endoscopists' measurements on 542 images (Dataset 3).

Results: The PolypM trained on Dataset 1 achieved an intersection over union (IoU) of 0.91 [95% confidence interval (CI): 0.89-0.93] for segmenting transparent caps, an IoU of 0.75 (95% CI: 0.71-0.79) for segmenting polyps, and an intraclass correlation coefficient (ICC) of 0.682 compared to the gold standard in Dataset 2. The PolypM also demonstrated comparable accuracy on Dataset 3. In the multicenter external validation, the PolypM outperformed endoscopists in both average absolute error and average relative error in determining polyp size ($P<0.05$) regardless of polyp morphology, pathological characteristics, or endoscopists' experience.

Conclusions: PolypM, an interpretable system for colonic polyp size estimation, was developed to mitigate measurement ambiguity and facilitate decision-making regarding surgical intervention and postoperative surveillance timing.

Key words: Artificial intelligence: Small polyps: Size measurement: Transparent caps:

Interpretable.

Key points

1. Superior Segmentation. The PolypM achieves high segmentation accuracy for both transparent caps (IoU 0.91) and polyps (IoU 0.79).
2. Reliable Size Measurement. The PolypM demonstrates stable polyp size measurement (ICC 0.770) against the gold standard in validation.
3. Outperforms Endoscopists. The PolypM shows significantly higher consistency (ICC 0.770 vs. 0.363) and lower error than endoscopists in size estimation.
4. Enhances Clinical Decisions. The PolypM significantly improves treatment decision accuracy across all endoscopist levels and polyp types.

1.Introduction

Colorectal polyps are intimately associated with the pathogenesis of colorectal carcinoma, making early intervention vital[1,2]. International guidelines have established a 10-mm threshold for post-polypectomy surveillance intervals[3]. Due to their high malignant potential, lesions ≥10 mm are unequivocally recommended for immediate endoscopic resection[4]. However, small polyps (<10 mm) account for 90% of detected lesions[5], yet their management remains debated, necessitating further research.

Guidelines recommend size-specific resection techniques to balance costs and risks[6]. Cold snare polypectomy is preferred for polyps ≤5 mm and sessile 6-9 mm polyps, while hot snare polypectomy is recommended for pedunculated polyps[6]. Accurate size measurement is critical for selecting the correct technique, enabling personalized treatment, and supporting strategies like "resect-and-discard" (forgoing histology for small polyps)[7]. Inaccurate measurement risks polyp misclassification, incorrect monitoring intervals, and missed advanced histology[8].

Current measurement methods are evolving. Endoscopist visual estimation is common but often overestimates size, especially without calibration[9]. Tools like forceps[10] are used but limited by distance and reliance on estimation. Calibration devices (e.g., needles[11], ruler snares[12]) reduce bias but are time-consuming.

Advances in artificial intelligence show promise. Research teams in Korea and Japan have developed vascular segmentation-based algorithms and accurate[13], automated real-time laser systems[14], respectively. However, their clinical adoption remains limited by operational latency and substantial hardware costs. Combining affordable reference tools with novel algorithms could enable faster, more accurate sizing.

Transparent caps, commonly used to improve polyp detection by flattening folds, offer a measurement opportunity[15]. Utilizing these caps, our study developed a novel system integrating them with artificial intelligence algorithms. This enables real-time polyp size estimation, with planned multicenter validation to facilitate future clinical application.

2. Methods

2.1. Datasets and Preprocessing

Three endoscopic colonic polyp image datasets were utilized. All images were acquired using Olympus endoscope systems. Dataset 1, retrospectively collected from Xiangya Hospital Central South University, served for model training and testing. Inclusion required white light images containing both polyps and transparent caps. Exclusion criteria encompassed: (1) caps with >25% outside the field of view; (2) polyps too distant from caps, filling the entire view, or exceeding the cap's inner ring diameter; and (3) poor-quality images (containing foam, biopsy forceps, or poor preparation). Patient-level separation ensured no image overlap between training and testing sets. Dataset 2, comprising images clipped from Xiangya Hospital (September 2022- April 2023), validated model performance in identifying cap inner rings and polyp boundaries, and assessing polyp size measurement accuracy. Dataset 3, a multicenter dataset retrospectively collected from Xiangya Hospital, Changsha Eighth Hospital, and Xiangtan Central Hospital (May - July 2023), tested the system's clinical polyp size measurement accuracy against endoscopist judgments. Images originated from 3 junior (YLH, CCQ, BL) and 3 senior (GHL, FJL, XWL) endoscopists.

2.2. Polyp Size Measurement System (PolypM)

The PolypM system employed two segmentation models based on the Part Relation Attention Network (PRA-Net), selected for its capability to capture global context and local features, handle variations in size, color, and texture, and maintain distinction despite unclear object-background boundaries[16]. Model 1 segmented transparent caps and identified the inner ring. Model 2 segmented polyps. The system workflow involved: Model 1 segmenting the cap, inverting this segmentation to isolate the inner area, and calculating the minimum circumscribed rectangle of this inner area, with its longer side representing the observable cap diameter. Subsequently, Model 2 segmented the polyp, and the minimum circumscribed rectangle for the polyp contour defined its long and short observable diameters. The actual polyp long diameter was determined via proportional conversion using the known inner cap diameter **(Figure 1)**.

2.3. Model Construction

Two doctoral students annotated caps and polyps in Datasets using the VGG Image Annotator. Two independent endoscopists reviewed annotations, with a third resolving discrepancies to establish gold standards. Dataset 1 images were split into training (90%) and validation (10%) sets. Training utilized the PyTorch 1.9.0+cu111 framework. Input images were resized to 352x352 pixels. Training employed a batch size of 16, a gradient clipping margin of 0.5 for stability, and the Adam optimizer with an initial learning rate of 0.0001 (decayed by a factor of 0.1 every 50 epochs). Training continued until errors stabilized over 10 consecutive rounds. Model 1 was trained on 1,450 images: Model 2 on 5,036 images.

2.4. Model Testing and Comparison

Dataset 2 was employed to evaluate segmentation performance for caps and polyps, and the system's accuracy in measuring polyp sizes. Dataset 3 compared the system's polyp size measurements against endoscopist assessments using digital calipers as the gold standard[17]. Both endoscopist measurements and model outputs were collected under blinding to minimize bias.

2.5. Evaluation Metrics

Segmentation performance was evaluated using Intersection over Union (IoU), defined as the intersection of the predicted range and the gold standard divided by their union. The mean IoU (mIoU) was calculated[18]. Polyp size measurement accuracy was assessed using absolute error (the absolute difference between the predicted value and the gold standard value) and relative error (absolute error divided by the gold standard value)[14].

2.6. Statistical Analysis

Continuous variables are presented as medians with interquartile ranges. Student's t-test or the Mann-Whitney test (for non-normally distributed data) analyzed differences. Sub-group analyses stratified by Boston Bowel Preparation Scale (BBPS) score, polyp morphology, and size were performed to evaluate PolypM's performance under varying conditions. Paired tests compared the measurement abilities of the system and endoscopists. Categorical variables are expressed as n (%) and analyzed using Chi-square tests. All P-values are two-tailed, with $P < 0.05$ indicating statistical significance. Analyses used SPSS version 26.0.

3. Results

3.1 System construction

The PolypM system was successfully constructed with three integrated computational modules: a transparent cap segmentation model (Model 1), a polyp segmentation model (Model 2), and a polyp size measurement algorithm.

3.2 Demographics

Dataset 1 comprised 6486 retrospectively collected endoscopic images from Xiangya Hospital (September 2018-September 2022). Dataset 2 included 675 polyps from 495 patients, including 140 patients with 2 polyps and 20 with 3 polyps, exhibiting median polyp size of 5.23 mm (IQR: 4.00-6.00). Dataset 3 contained 542 polyps from 388 multicenter patients, with 82 patients having 2 polyps and 36 having 3 polyps. Both validation datasets demonstrated comparable distributions for age, sex, Paris classification morphology, and anatomical location (all P>0.05). The proportion of polyps with a size ≤5 mm was higher in both datasets, at 61.48% and 60.15%, respectively **(Table 1)**.

3.3 Performance of PolypM

In Dataset 2 validation, Model 1 achieved exceptional cap segmentation performance with IoU 0.91 (95%CI: 0.89-0.93) and sensitivity 98.94%. Model 2 attained polyp segmentation IoU 0.79 (95%CI: 0.75-0.82) with 87.71% accuracy. We also evaluated the performance on dataset 3 **(Table 2)**.

For size measurement, the system showed an intraclass correlation coefficient (ICC) of 0.682 compared to the gold standard, while Dataset 3 testing demonstrated improved reliability (ICC=0.770). At the clinically critical 5mm size threshold, misclassification occurred in 130/675 (19.3%) and 94/542 (17.3%) polyps in datasets 2 and 3 respectively (Table 3). Subgroup analysis stratified by BBPS scores revealed bowel preparation cleanliness exerted negligible influence on measurement accuracy **(Table 3)**.

3.4 Comparison with endoscopists

When benchmarked against endoscopists in Dataset 3, PolypM demonstrated significantly superior agreement with gold standard measurements (ICC=0.770) compared to visual estimation (ICC=0.363; P<0.001). The system's absolute error and relative error were significantly lower than endoscopists' (both P<0.001) **(Table 4)**. This advantage persisted across all polyp morphologies and pathological subtypes (P<0.05). Post-hoc resection analysis confirmed the system improved correct surgical method selection (P<0.001).

Experience-stratified analysis revealed junior endoscopists had non-significantly higher median absolute error than seniors [2.28mm (IQR:1.00-3.00) vs 2.01mm (IQR:1.00-3.00); P=0.185]. Crucially, PolypM significantly outperformed manual judgment in both junior and senior groups (P<0.05) and enhanced therapeutic decision accuracy regardless of operator experience **(Table 5)**.

4. Discussion

The PolypM system achieves a paradigm shift in diminutive polyp measurement, transforming subjective, experience-dependent assessment into objective quantification through strategic integration of hardware and algorithmic components. While conventional visual estimation remains clinically prevalent, its inherent subjectivity causes significant variability that may compromise therapeutic decisions[19,20]. This multicenter validation study confirms the system's ability to achieve high measurement consistency across diverse morphological subtypes and operator experience levels for polyps under 10 mm, a breakthrough resulting from synergistic dual-innovation mechanisms.

Compared to existing artificial Intelligence (AI)-assisted sizing systems, PolypM demonstrates distinct superior real-time applicability and cost-effectiveness [21-23]. Virtual scale endoscope and laser-based approaches require dedicated image processing pipelines for scale calibration, prolonging estimation time, and rely on specialized attachments that are costly to procure and deploy. In contrast, PolypM utilizes ubiquitous transparent caps as intrinsic calibration references, eliminating additional hardware needs. Its lightweight algorithm runs efficiently on standard endoscopy processors, enabling real-time analysis without disrupting clinical workflows. Crucially, although current AI systems struggle with irregular mucosal surfaces and perform poorly in hemorrhagic fields, our cap-based approach maintains robustness across varying bowel preparation qualities. This hardware-software synergy embodies a pragmatic fusion of artificial intelligence with routine clinical apparatus, enabling scalable deployment without procedural modifications or ancillary investments while proving particularly valuable in resource-constrained settings.

The primary technical advancement involves spatial recalibration enabled by the transparent cap. As a routine endoscopic accessory, the cap deforms mucosal folds through physical contact[24], establishing a stabilized environment around lesions. When the cap's planar surface perpendicularly contacts the intestinal wall, its inner ring forms a geometrically invariant reference circle that minimizes the distance between the measurement benchmark and the polyp. By contrast, conventional biopsy forceps or rulers must be placed at a distance from lesions, where perspective magnification induces substantial size overestimation, particularly problematic for polyps under 5 mm[25]. Crucially, the cap's optical transparency permits direct capture of scale-polyp edge spatial relationships, providing undistorted topological data for algorithmic processing along with procedural accessibility and reduced invasivity.

Complementing this hardware innovation, the algorithm mitigates cognitive biases. Using a cascaded analytical architecture, the system first segments the cap's inner ring using PRA-Net, establishing a polar coordinate system that converts non-linear Cartesian distortions into radial linear scales. Subsequently, dynamic adaptive thresholding identifies polyp contour critical points, enabling subpixel-level edge-scale alignment. This computational process eliminates characteristic human visual limitations: specifically, termination digit preference[9,23] (e.g., rounding 4.1 mm or 4.9 mm to 5.0 mm), and experience-dependent interobserver variation[26]. The combined geometric benchmark and intelligent parsing ensure precision independent of morphological complexity or operator skill disparity.

In this study, this synergistic mechanism from PolypM delivers multidimensional clinical value through therapeutic decision refinement via millimeter-accurate classification that corrects erroneous clinical strategies. The procedural safety is enhanced by reduced tissue injury risk versus biopsy forceps[24] and by improved patient tolerance[9]. Meanwhile, technical accessibility is achieved through standardized operation that provides diagnostic-grade precision without capital expenditure.

However, current limitations warrant deliberate consideration. First, the cap-wall contact requirements may prolong procedure duration, necessitating real-world prospective quantification. Second, architectural constraints restrict application to polyps under 10 mm, demanding scaling solutions for larger lesions. And finally, single-center training data (Dataset 1) requires prospective validation to confirm generalizability across endoscopic systems and populations.

Collectively, the study establishes a novel paradigm for objective, experience-independent polyp measurement that resolves critical limitations in contemporary endoscopic practice through strategic computational integration with established clinical instrumentation.

Key Points Table

What was previously known about the topic of the study

Accurate polyp size measurement is crucial for optimal clinical decisions, but the existing methods were unreliable, and more accurate solutions were either impractical or too expensive for widespread use. Transparent caps were known aids for detection, but not specifically for measurement. This gap in practical, accurate, and accessible real-time measurement technology formed the basis for the new study combining AI with transparent caps.

What the study contributes

This study pioneered the integration of transparent caps with AI algorithms to develop PolypM, a real-time polyp measurement system. In validation, it achieved an intraclass correlation coefficient (ICC) of 0.77 significantly outperforming endoscopists. It reduced size errors for polyps <10 mm and improved surgical selection accuracy, regardless of operator experience. This offers a cost-effective clinical solution for small polyp management.

How the results will influence clinical practice

This study significantly enhances the measurement accuracy of polyps <10 mm through PolypM's objective assessment, advancing guideline-based standardization in surgical technique selection and surveillance intervals. By eliminating operator-dependent variability and leveraging cost-effective transparent cap adaptation, the system transforms small polyp management from subjective evaluation into routine precision medicine practice.

Acknowledgements: Not applicable

Funding

This work was supported by the National Natural Science Foundation of Hunan Province (Grant No.2023JJ30947) and the National Natural Science Foundation of Changsha City (Grant No. kq2208374).

Reference

1 Patel, S. G., Karlitz, J. J., Yen, T., Lieu, C. H. & Boland, C. R. The rising tide of early-onset colorectal cancer: a comprehensive review of epidemiology, clinical features, biology, risk factors, prevention, and early detection. Lancet Gastroenterol Hepatol 7, 262-274, doi:10.1016/S2468-1253(21)00426-X (2022).

2 Zauber, A. G. et al. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. N Engl J Med 366, 687-696, doi:10.1056/NEJMoa1100370 (2012).

3 Abu-Freha, N. et al. Post-polypectomy surveillance colonoscopy: Comparison of the updated guidelines. United European Gastroenterol J 9, 681-687, doi:10.1002/ueg2.12106 (2021).

4 Hazewinkel, Y. et al. Prevalence of serrated polyps and association with synchronous advanced neoplasia in screening colonoscopy. Endoscopy 46, 219-224, doi:10.1055/s-0033-1358800 (2014).

5 Pickhardt, P. J. & Kim, D. H. Colorectal cancer screening with CT colonography: key concepts regarding polyp prevalence, size, histology, morphology, and natural history. AJR Am J Roentgenol 193, 40-46, doi:10.2214/AJR.08.1709 (2009).

6 Ferlitsch, M. et al. Colorectal polypectomy and endoscopic mucosal resection (EMR): European Society of Gastrointestinal Endoscopy (ESGE) Clinical

Guideline. Endoscopy 49, 270-297, doi:10.1055/s-0043-102569 (2017).
7 Denis, B. et al. Some diminutive colorectal polyps can be removed and discarded without pathological examination. Endoscopy 43, 81-86, doi:10.1055/s-0030-1255952 (2011).
8 Vleugels, J. L. A. et al. Diminutive Polyps With Advanced Histologic Features Do Not Increase Risk for Metachronous Advanced Colon Neoplasia. Gastroenterology 156, 623-634 e623, doi:10.1053/j.gastro.2018.10.050 (2019).
9 Sakata, S., Klein, K., Stevenson, A. R. L. & Hewett, D. G. Measurement Bias of Polyp Size at Colonoscopy. Dis Colon Rectum 60, 987-991, doi:10.1097/DCR.0000000000000859 (2017).
10 Rex, D. K. & Rabinovitz, R. Variable interpretation of polyp size by using open forceps by experienced colonoscopists. Gastrointest Endosc 79, 402-407, doi:10.1016/j.gie.2013.08.030 (2014).
11 Sakata, S., McIvor, F., Klein, K., Stevenson, A. R. L. & Hewett, D. G. Measurement of polyp size at colonoscopy: a proof-of-concept simulation study to address technology bias. Gut 67, 206-208, doi:10.1136/gutjnl-2016-312915 (2018).
12 Kaz, A. M., Anwar, A., O'Neill, D. R. & Dominitz, J. A. Use of a novel polyp "ruler snare" improves estimation of colon polyp size. Gastrointest Endosc 83, 812-816, doi:10.1016/j.gie.2015.08.082 (2016).
13 Kwak, M. S., Cha, J. M., Jeon, J. W., Yoon, J. Y. & Park, J. W. Artificial intelligence-based measurement outperforms current methods for colorectal polyp size measurement. Dig Endosc 34, 1188-1195, doi:10.1111/den.14318 (2022).
14 Shimoda, R. et al. Estimating colorectal polyp size with a virtual scale endoscope and visual estimation during colonoscopy: Prospective, preliminary comparison of accuracy. Dig Endosc 34, 1471-1477, doi:10.1111/den.14351 (2022).
15 Mir, F. A. et al. Cap-assisted colonoscopy versus standard colonoscopy: is the

cap beneficial? A meta-analysis of randomized controlled trials. Ann Gastroenterol 30, 640-648, doi:10.20524/aog.2017.0180 (2017).

16 Liu, Z. et al. PRA-Net: Part-and-Relation Attention Network for depression recognition from facial expression. Comput Biol Med 157, 106589, doi:10.1016/j.compbiomed.2023.106589 (2023).

17 Sudarevic, B. et al. Artificial intelligence-based polyp size measurement in gastrointestinal endoscopy using the auxiliary waterjet as a reference. Endoscopy 55, 871-876, doi:10.1055/a-2077-7398 (2023).

18 Wang, J. et al. An interpretable artificial intelligence system for detecting risk factors of gastroesophageal variceal bleeding. NPJ Digit Med 5, 183, doi:10.1038/s41746-022-00729-z (2022).

19 Hassan, C. et al. Post-polypectomy colonoscopy surveillance: European Society of Gastrointestinal Endoscopy (ESGE) Guideline - Update 2020. Endoscopy 52, 687-700, doi:10.1055/a-1185-3109 (2020).

20 Leng, Q. & Jin, H. Y. Measurement system that improves the accuracy of polyp size determined at colonoscopy. World J Gastroenterol 21, 2178-2182, doi:10.3748/wjg.v21.i7.2178 (2015).

21 Haumesser, C. et al. Comparing size measurements of simulated colorectal polyp size and morphology groups when using a virtual scale endoscope or visual size estimation: Blinded randomized controlled trial. Dig Endosc 35, 638-644, doi:10.1111/den.14498 (2023).

22 Djinbachian, R. et al. Comparing size measurement of colorectal polyps using a novel virtual scale endoscope, endoscopic ruler or forceps: A preclinical randomized trial. Endosc Int Open 11, E128-E135, doi:10.1055/a-2005-7548 (2023).

23 Sharma, P., Frye, J. & Frizelle, F. Accuracy of visual prediction of pathology of colorectal polyps: how accurate are we? ANZ J Surg 84, 365-370, doi:10.1111/ans.12366 (2014).

24 Rex, D. K. et al. Quality indicators for colonoscopy. Gastrointest Endosc 81, 31-53, doi:10.1016/j.gie.2014.07.058 (2015).

25 Shaw, M. J. & Shaukat, A. Does polyp size scatter matter? Gastrointest Endosc 83, 209-211, doi:10.1016/j.gie.2015.08.060 (2016).

26 Eichenseer, P. J., Dhanekula, R., Jakate, S., Mobarhan, S. & Melson, J. E. Endoscopic mis-sizing of polyps changes colorectal cancer surveillance recommendations. Dis Colon Rectum 56, 315-321, doi:10.1097/DCR.0b013e31826dd138 (2013).

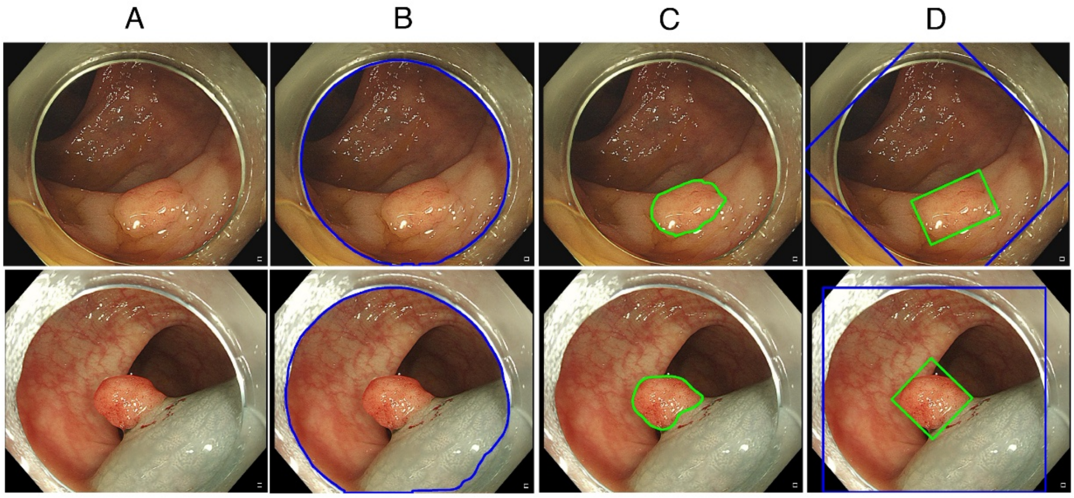


Figure 1. Representative endoscopic images of how the PolypM performs. A. The first column shows the original endoscopic images. B. The second column shows how the Model 1 segment and determine the transparent caps (Blue box). C. The third column shows how the Model 2 segment and determine the polyps (Green box). D. The last column shows the PolypM generated a minimum circumscribed rectangle of transparent caps and polyps to measure the size of polyps.

Table 1 The characteristics of Patients and polyps

Characteristics	Dataset 2	Dataset 3	P value
Age, years, mean±SD	57.00±10.12	56.03±12.95	0.494
Male, n (%)	334 (67.47)	249 (64.18)	0.142
Polyps, n (%)	675 (55.46)	542 (44.54)	
Location, n (%)			0.438
Cecum/ascending colon	150 (22.22)	90 (16.61)	
Transverse colon	145 (21.48)	104 (19.19)	
Descending/sigmoid colon	300 (44.44)	272 (50.18)	
Rectum	80 (11.85)	76 (14.02)	
Paris class, n (%)			0.110
I	510 (75.56)	446 (82.29)	
II	165 (24.44)	96 (17.71)	
Polyp size, mean (IQR)	5.23 (4.00, 6.00)	5.37 (4.00, 7.00)	
Polyp size, n (%)			0.796
≤ 5mm	415 (61.48)	326 (60.15)	
> 5mm	260 (38.52)	216 (39.85)	
BBPS score of each location, n (%)			0.526
0-1	100 (14.81)	68 (12.55)	
2-3	575 (85.19)	474 (87.45)	

Abbreviation: SD, standard deviation; IQR, interquartile range; BBPS, Boston bowel preparation scale

Table 2. The performance of Model 1 and Model 2 validated in dataset 2 and 3

	Dataset 2 (Validation dataset)	Dataset 3 (Multicenter dataset)
Model 1 (Caps segmentation model)		
mIoU (95% CI)	0.91 (0.89, 0.93)	0.93 (0.91, 0.94)
Sensitivity, % (Per image, IoU > 0.5)	98.94	98.89
Model 2 (Polyps segmentation model)		
mIoU (95% CI)	0.79 (0.75, 0.82)	0.81 (0.78, 0.84)
Sensitivity, % (Per image, IoU > 0.5)	87.71	92.22

Abbreviation: IoU, intersection over union; CI, confidence interval

Table 3. Polyp size measurement performance for dataset 2 and 3.

Outcomes	Dataset 2	Dataset 3
MAE, mean (IQR)	1.21 (1.00, 2.00)	1.01 (0.00, 2.00)
MRE, % (IQR)	23.77 (14.29, 33.33)	18.39 (0.00, 28.57)
ICC (95%CI)	0.682 (0.548-0.784)	0.770 (0.658-0.851)
Polyp size		
≤ 5mm		
MAE (IQR)	1.04 (1.00, 1.00)	0.74 (0.00, 1.00)
MRE, % (IQR)	25.36 (20.00, 40.00)	17.72 (0.00, 25.00)
≤5 mm misclassified as >5 mm, n (%)	5(1.20)	2 (0.61)
> 5mm		
MAE (IQR)	1.50 (1.00, 2.00)	1.42 (1.00, 2.00)

MRE, % (IQR)	21.24 (12.95, 33.33)	19.40 (12.95, 28.57)
> 5 mm misclassified as ≤5 mm, n (%)	125 (48.08)	92 (42.59)
BBPS score		
≤ 1		
MAE (IQR)	1.15 (1.00, 1.00)	0.74 (0.00, 1.00)
MRE, % (IQR)	23.49 (17.50, 33.33)	15.28 (0.00, 25.00)
> 1		
MAE (IQR)	1.23 (1.00, 2.00)	1.05 (0.00, 2.00)
MRE, % (IQR)	23.82 (14.29, 33.33)	18.84 (0.00, 28.57)

Abbreviation: CI, confidence interval; MAE, mean absolute error; MRE, mean relative error; IQR, interquartile range; ICC, intraclass correlation coefficient; BBPS, Boston bowel preparation scale

Table 4. Comparison performance between PolypM and endoscopists.

Outcomes	**PolypM**	**Endoscopists**	***P* value**
MAE **(IQR)**	1.01 (0.00, 2.00)	2.15 (1.00, 3.00)	<0.0001
MRE**, % (IQR)**	18.39 (0.00, 28.57)	47.60 (16.67, 66.67)	<0.0001
ICC (95%CI)	0.770 (0.658-0.851)	0.363 (0.168- 0.533)	
Polyp size			
≤ 5mm			
MAE (IQR)	0.74 (0.00, 1.00)	2.29 (1.00, 3.00)	<0.0001
MRE, % (IQR)	17.72 (0.00, 25.00)	61.50 (20.00, 100.00)	<0.0001
> 5mm			
MAE (IQR)	1.42 (1.00, 2.00)	1.94 (1.00, 3.00)	0.002

MRE, % (IQR)	19.40(12.95, 28.57)	26.61 (14.29, 42.86)	0.002
Paris class			
I			
MAE (IQR)	1.09 (0.00, 2.00)	2.17 (1.00, 3.00)	<0.0001
MRE, % (IQR)	19.70 (0.00, 28.57)	46.54 (16.67, 66.67)	<0.0001
II			
MAE (IQR)	0.67 (0.00, 1.00)	2.06 (1.00, 3.00)	<0.0001
MRE, % (IQR)	12.33 (0.00, 25.00)	52.52 (17.50, 70.24)	<0.0001
Pathology			
Hyperplastic polyps			
MAE (IQR)	0.92 (0.00, 2.00)	2.42 (0.25, 4.00)	0.001
MRE, % (IQR)	17.28 (0.00, 33.33)	52.39 (3.57, 93.75)	<0.0001
Adenomatous polyps			
MAE (IQR)	1.03 (0.00, 2.00)	2.22 (1.00, 3.00)	<0.0001
MRE, % (IQR)	18.56 (0.00, 28.57)	46.86 (16.67, 66.67)	<0.0001
Correct rate of surgical method selection, n (%)	448 (82.7)	324 (59.8)	<0.0001

Abbreviation: CI, confidence interval; MAE, mean absolute error; MRE, mean relative error; IQR, interquartile range; ICC, intraclass correlation coefficient

Table 5. Subgroup analysis of endoscopists' level

	Junior group (n=282)		***P* value**	**Senior group (n=260)**		***P* value**
	PolypM	**Endoscopists**		**PolypM**	**Endoscopists**	
MAE (IQR)	0.93 (0.00, 2.00)	2.28 (1.00, 3.00)	<0.0001	1.07 (0.00, 2.00)	2.01 (1.00, 3.00)	<0.0001
MRE, % (IQR)	16.71 (0.00, 25.00)	49.76 (20.00, 66.67)	<0.0001	19.54 (0.00, 28.57)	45.26 (14.29, 66.67)	<0.0001
Correct rate of surgical method selection, n (%)	234 (83.0)	151 (53.5)	<0.0001	214 (82.3)	173 (66.5)	<0.0001

Abbreviation: MAE, mean absolute error; MRE, mean relative error; IQR, interquartile range