scientific data

Check for updates

OPEN

DATA DESCRIPTOR

Polyp-Size: A Precise Endoscopic Dataset for AI-Driven Polyp Sizing

Yiming Song[1,7], Sijia Du[2,7], Ruilan Wang[3], Fei Liu[4], Xiaolu Lin[5], Jinnan Chen[1], Zeyu Li[1], Zhao Li[1], Liuyi Yang[1], Zhengjie Zhang[2], Hao Yan[6], Qingwei Zhang[1,8] ✉, Dahong Qian[2,8] ✉ & Xiaobo Li[1,8] ✉

Colorectal cancer often arises from precancerous polyps, where accurate size assessment is vital for clinical decisions but challenged by subjective methods. While artificial intelligence (AI) has shown promise in improving the accuracy of polyp size estimation, its development depends on large, meticulously annotated datasets. We present Polyp-Size, a dataset of 42 high-resolution white-light colonoscopy videos with polyp sizes precisely measured post-resection using vernier calipers to submillimeter precision. Unlike existing datasets primarily focused on polyp detection or segmentation, Polyp-Size offers validated size annotations, diverse polyp features (Paris classification, anatomical location and histological type), and standardized video formats, enabling robust AI models for size estimation. By making this resource publicly available, we aim to foster research collaboration and innovation in automated polyp measurement to ultimately improve clinical practice.

Background & Summary

Colorectal cancer (CRC) often stems from precancerous polyps[1,2], and early removal of precancerous lesions via colonoscopy could significantly reduce its incidence and mortality[3,4]. Accurate polyp size measurement is critical for assessing malignancy risk and guiding clinical decisions such as selecting the appropriate resection technique and determining surveillance intervals[5–10]. In clinical practice, however, visual assessment-based polyp size estimation is often subjective and imprecise[11]. Reference objects with fixed dimensions, like open biopsy forceps, are infrequently positioned adjacent to the lesion for size measurement. Other methods, including post-resection measurements with linear probes or rulers, are infrequently employed due to their impracticality[11]. Accurate polyp size estimation presents significant challenges for both endoscopists and pathologists. Inconsistent size estimation arises from the following factors[11–13]:

- Interobserver variability: Visual estimation by endoscopists often leads to significant overestimation or underestimation of polyp size due to subjective judgment.
- Endoscopic magnification and distortion: Wide-angle lenses at the endoscope tip magnify and distort the live view, leading to overestimation of polyp dimensions.
- Post-resection tissue shrinkage: Fixation processes cause tissue to shrink, reducing measured size and causing underestimation.
- Piecemeal removal: Incomplete or piecemeal excision of polyps results in underestimation of true size.
- Variability in measurement tools: Variability in tools introduces discrepancies in size measurements.

These inconsistencies can lead to overestimation, prompting unnecessary surveillance, or underestimation, delaying critical diagnoses and increasing the risk of progression to CRC[14]. These inconsistencies might impact

[1]Division of Gastroenterology and Hepatology, Shanghai Institute of Digestive Disease, NHC Key Laboratory of Digestive Diseases, Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China. [2]School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China. [3]Department of Gastroenterology, Armed Police Forces Hospital of Sichuan, Leshan, Sichuan Province, China. [4]Departmant of Gastroenterology, Nine Division Hospital of Xinjiang Production and Construction Corps, Tacheng Xinjiang Uygur Autonomous Region, Tacheng, China. [5]Department of Digestive Endoscopy Center, Fujian Provincial Hospital, Shengli Clinical Medical College of Fujian Medical University, Fuzhou, Fujian, China. [6]The Second Clinical Medical College, Harbin Medical University, Harbin, 150081, China. [7]These authors contributed equally: Yiming Song, Sijia Du. [8]These authors jointly supervised this work: Qingwei Zhang, Dahong Qian, Xiaobo Li ✉e-mail: zhangqingweif@hotmail.com; dahong.qian@sjtu.edu.cn; lxb_1969@163.com

treatment decisions and highlight the urgent need for standardized, reliable, and reproducible methods of polyp size estimation.

Advances in artificial intelligence (AI) have shown great promise in enhancing endoscopic polyp assessments through multiple distinct applications, including polyp detection, segmentation, classification, and characterization[15–17]. AI systems for polyp detection focus on identifying the presence of lesions, increasing detection rates[18–24]. Recent studies have also explored ways to improve their real-time applicability[25,26]. Separate AI approaches for polyp segmentation delineate polyp boundaries, creating accurate outlines of lesion areas that can improve dataset development and subsequent detection systems[27,28]. More recently, AI systems for polyp size estimation have been developed to address the specific challenge of dimensional measurement, outperforming traditional methods[14,29,30]. However, a robust AI system typically requires large amounts of accurately annotated data for model training, testing, and validation. Recent large-scale surveys of biomedical image analysis competitions have highlighted the importance of standardized datasets and evaluation frameworks in driving progress in this area[31]. One possible solution is the use of semi-automated annotation frameworks, which can reduce expert workload while maintaining annotation quality, particularly in endoscopic imaging tasks[32]. While many datasets focus on polyp segmentation and detection[33–45], few provide real-world size measurements validated through direct post-resection measurements. As shown in Table 1, only three datasets contain polyp size information, yet both include size as secondary metadata with millimeter-level precision using undocumented measurement methods, and primarily target detection or classification tasks rather than addressing the unique computational challenges of size estimation.

To address this gap, we introduce the Polyp-Size dataset, the first of its kind featuring 42 high-resolution white-light endoscopic videos of polyps, each annotated with high-precision measurements (0.01 mm) using vernier calipers following a standardized post-resection protocol. This dataset captures polyps from multiple viewing angles and is supplemented with comprehensive clinical metadata including polyp Paris classification, anatomical locations, and pathological findings.

The primary goals of the Polyp-Size dataset are as follows:

- To establish a high-quality white-light endoscopic video collection with validated post-resection measurements using vernier calipers as gold standard.
- To facilitate the development and evaluation of AI systems for accurate polyp size estimation.
- To promote research collaboration and advance AI-driven approaches in management of precancerous lesions.
- To promote standardization and efficiency in CRC prevention strategies globally through improved polyp measurement techniques.

This paper first introduces the research background and importance of polyp size estimation, then details the data acquisition, processing, and metadata collection, followed by a presentation of the data record organization, technical validation to assess dataset applicability, and finally discusses the dataset limitations and application prospects.

Methods

Fig. 1 illustrates the detailed workflow for constructing the Polyp-Size dataset. The process began with data acquisition at Renji Hospital, Shanghai, where colonoscopy videos were captured, and polyps were measured with precision following complete resection. Of the 64 initial recordings, 42 high-quality videos were selected by expert endoscopists. These recordings underwent a comprehensive standardization process, including the extraction of white-light endoscopy segments, removal of the left-side panel and peripheral margins, and format unification. The final dataset is accompanied by comprehensive metadata detailing polyp measurements and characteristics for each video segment.

Data source and collection. The dataset originates from the Digestive Endoscopy Center of Renji Hospital, affiliated with Shanghai Jiao Tong University School of Medicine. Data were collected between November 7, 2023, and April 25, 2024, from 47 patients aged over 18.

Ethics approval and privacy protection. All participants provided written informed consent prior to the colonoscopy procedure, informing them that videos of polyps identified during the procedure would be recorded for research purposes. Participants were also informed that following removal, polyps would be precisely measured and these measurement data would be included in the public dataset. Participants were assured that their personal information (such as name, ID, date, etc.) would be anonymized and that all collected data, including both video recordings and polyp measurements, would be publicly released after collection. The study adhered to the principles of the Declaration of Helsinki, and the data collection, annotation, and dissemination processes were approved by the Ethics Committee of Renji Hospital, with approval number KY2023-002-B. All personally identifiable information was removed to ensure patient anonymity and compliance with ethical standards.

Colonoscopy procedure protocol. The procedures were performed under intravenous anesthesia using an Olympus endoscopy system (CV-290 main unit paired with CF-HQ290I/CF-H290I colonoscopes; Olympus Medical Corporation, Tokyo, Japan). Patients were positioned in the lateral decubitus position. The colonoscope was advanced to the terminal ileum and then slowly withdrawn, systematically examining the entire colon, including the terminal ileum, ileocecal valve, cecum, ascending colon, hepatic flexure, transverse colon, splenic

Dataset	No. of polyps	Core Dataset Elements	Available link	Includes quantitative polyp size data (Yes/No)	Application
CVC-ColonDB[33]	380 images	380 polyp images along with binary masks as ground truth	http://vi.cvc.uab.es/colon-qa/cvccolondb/	No	Polyp detection
CVC-ClinicDB[34]	612 images	612 images from 31 colonoscopy sequences	https://polyp.grand-challenge.org/CVCClinicDB/	No	Polyp segmentation and detection
CVC-EndoScene Still[35]	912 images	a combination of CVC-ColonDB and CVC-ClinicDB.	Not available now	No	Polyp segmentation
ASU-Mayo[33]	5,200 images	19,400 images with 5,200 polyps and 14,200 without polyps, with binary masks as ground truth	https://polyp.grand-challenge.org/AsuMayo/	No	Polyp detection
Kvasir-Seg[36]	1,000 images	1,000 polyp images and its corresponding masks	https://datasets.simula.no/kvasir-seg/	No	Polyp segmentation, detection, localization, and classification
PICCOLO[37]	3,433 images	3,433 images, 76 different lesions with corresponding clinical metadata	https://www.biobancovasco.org/en/Sample-and-data-catalog/Databases/PD178-PICCOLO-EN.html	Yes	Polyp classification
PolypGen[38]	3,762 images	6,282 images with 3,762 polyps and 2,520 without polyps	https://www.synapse.org/Synapse:syn26376615/wiki/613312	No	Polyp detection and segmentation
CP-CHILD[39]	1,400 images	9,500 images with 1,400 polyps and 8,100 without polyps	https://figshare.com/articles/dataset/CP-CHILD_zip/12554042	No	Polyp detection
PIBAdb[40]	31,400 images	45,400 images with 31,400 polyps and 14,000 without polyps, each polyp images with corresponding histological report	https://www.iisgaliciasur.es/home/biobanco/colorectal-polyp-image-cohort-pibadb/?lang=en	No	Polyp classification
SUN Colonoscopy Video DB[41]	49,136 images	158,690 images with 49,136 polyps and 109,554 without polyps	http://sundatabase.org/	Yes	Polyp detection
LDPolyp video DB[42]	405,284 images	160 annotated videos with 33,884 polyps and 6,382 without polyps, 103 unannotated videos with 371,400 polyps and 490,000 without polyps	https://github.com/dashishi/LDPolypVideo-Benchmark	No	Polyp detection
BKAI-IGH NeoPolyp DB[40]	1,000 images	The database has two versions. BKAI-IGH NeoPolyp-Small: 1200 images with neoplastic (red) and non-neoplastic (green); NeoPolyp-Large: 7500 images, includes 'undefined' polyps (yellow).	https://www.kaggle.com/c/bkai-igh-neopolyp/	No	Polyp segmentation and identification
Endotest[43]	48,641 frames	253,754 images with 48,641 polyps and 205,113 without polyps	https://www.ukw.de/research/inexen/ai-for-polyp-detection/	No	Polyp detection
GLRC[44]	76 videos	76 videos with ground truth of histopathology, endoscopist inspection, and camera calibration	http://www.depeca.uah.es/colonoscopy_dataset/	No	Polyp classification and detection
ETIS-Larib[45]	196 images	196 images with 1 polyp each; 44 unique polyps from 34 sequences	Not available now	No	Polyp detection
ERCPMP[65]	796 images and 21 videos	796 images and 21 videos from 191 patients with colorectal polyps, with morphological and histopathological labels	https://databiox.com/	No	Polyp detection, classification and segmentation
REAL-Colon[66]	44 videos	44 videos with 113 histologically verified polyps, annotated frame-by-frame with corresponding clinical and size data	https://plus.figshare.com/articles/media/REAL-colon_dataset/22202866	Yes	Polyp detection, classification, and size estimation
Polyp-Size~ours	42 videos	42 videos with one polyp each, and accurate polyp size (in submillimeters)	https://doi.org/10.6084/m9.figshare.28030115	Yes	Polyp size estimation

Table 1. Comprehensive overview of existing polyp datasets, detailing the number of images and videos, core dataset elements, access links, and intended applications.

flexure, descending colon, sigmoid colon, rectum, and anal canal. Upon polyp detection, endoscopic video recordings were initiated to document the lesions *in situ*. Polypectomy with complete resection was performed according to standard clinical practice. Resected specimens were placed on sterile gauze, unfolded to maintain their original size, and trimmed to remove excess tissue. Each polyp was measured independently by three different endoscopists who had received standardized training on digital vernier caliper use (accurate to 0.01 mm). For each polyp, the maximum diameter was recorded under natural, uncompressed conditions, and the largest of the three measurements was selected as the final value, consistent with clinical risk assessment practice. Interobserver reliability analysis demonstrated exceptional consistency (ICC = 0.987, 95% CI: 0.979-0.993), confirming the objectivity and precision of our measurement protocol. The specimens were then fixed in formalin for histopathological evaluation. While videos were initiated upon polyp detection, Boston Bowel Preparation Scale (BBPS) scores had already been assigned to each colonic segment during the standard withdrawal phase of colonoscopy and were documented to facilitate subsequent quality assessment of polyp videos. After completing the colonoscopy, the videos were exported in MP4 format.

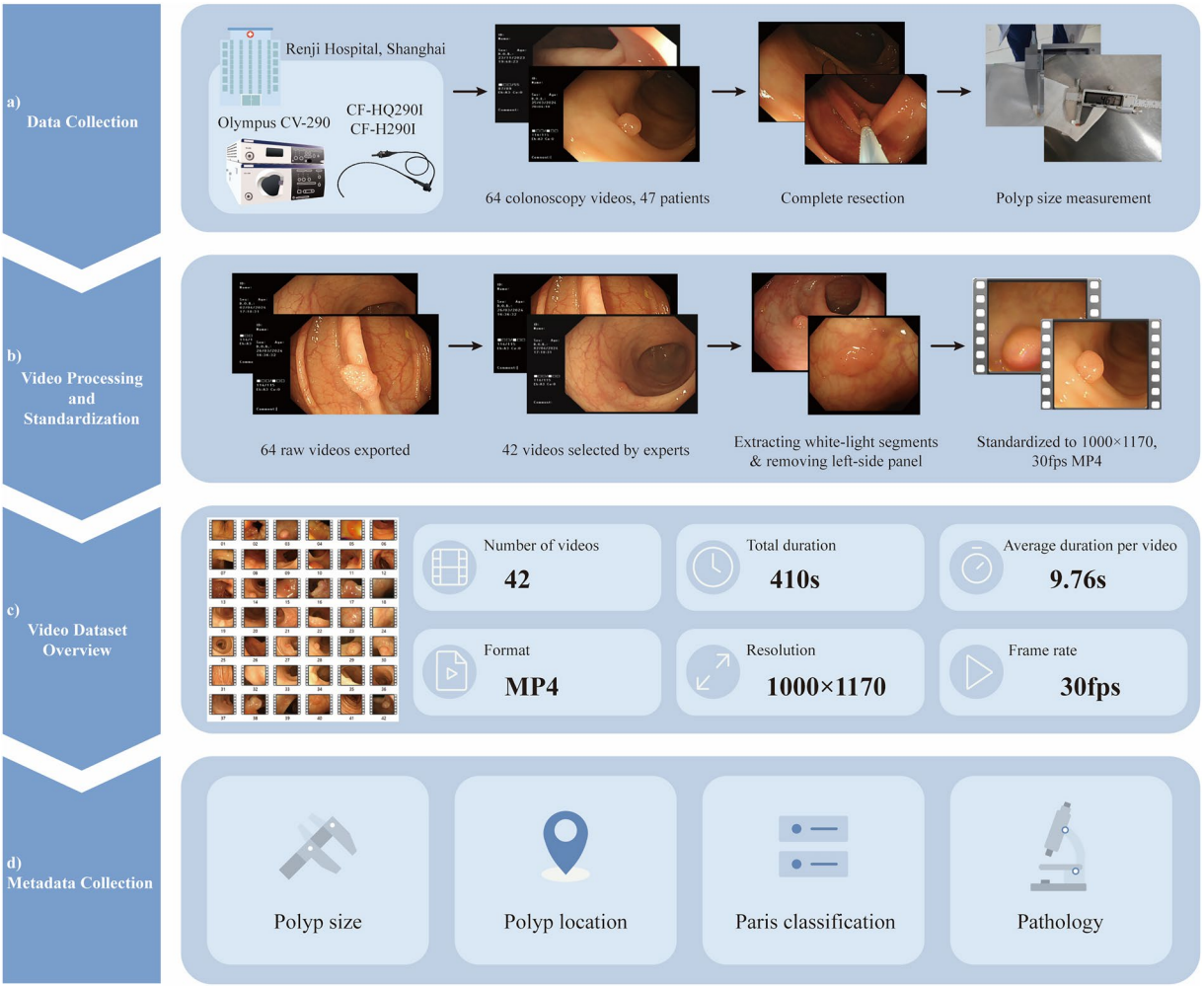

Fig. 1 Workflow illustrating the process for generating the Polyp-Size dataset.

Video processing and standardization. Videos were exported with only procedure time and colonoscope model information retained in the left panel, excluding all patient identifiers. Video processing involved stringent quality control. For quality assurance purposes, two experienced endoscopists, each with over 10 years of experience, independently assessed the recordings based on predefined criteria:

- Polyp visualization maintained in standard white-light mode for at least 3 seconds.
- Adequate bowel preparation in the visualized segment (corresponding to a BBPS score ≥ 2 for the specific colonic segment).
- Absence of excessive glare, blurring, foam, or instability that would interfere with proper polyp assessment.
- Clear and complete visualization of the colonic polyps (entire polyp margin visible).

This assessment was conducted to select high-quality videos. Of the 64 recorded videos, the first endoscopist selected 44 videos meeting the quality criteria, while the second endoscopist selected 43 videos. Taking the intersection of these two independent assessments, the final dataset consisted of 42 videos meeting all these criteria, selected by both reviewers. The inter-observer agreement was excellent, with a Cohen's Kappa coefficient of 0.89 ($p < 0.001$), demonstrating high consistency in applying our quality criteria.

These recordings underwent a comprehensive standardization process: endoscopists extracted video segments in standard white-light endoscopy mode (excluding narrow-band imaging and near focus modes). Then the videos were saved frame-by-frame at a rate of 30 frames per second (fps) in image format. The frames were then processed by removing the left-side panel along with peripheral black margins to retain only the effective endoscopic field of view. The image size was standardized to a fixed dimension of 1000×1170 (height $\times$ width). The processed images were then sequentially reassembled into a video in chronological order with the same fps as the original video. Fig. 2 illustrates the detailed workflow of polyp video export, selection, and processing using a flowchart. The standardized videos had a cumulative duration of 410 seconds, averaging 9.76 seconds per video, with a frame rate of 30 frames per second. Fig. 3 shows representative polyp video frames and their corresponding measurements.

Metadata collection and organization. For each video in the final dataset, comprehensive metadata was systematically documented to support clinical research and AI model development. The metadata was organized into three main categories:

1. Patient information: Demographic data including gender and age.
2. Equipment specifications: Details about the endoscopic system used.
3. Clinical data: Polyp characteristics and measurements of each polyp, including:

- **Polyp size**: Measured using calibrated vernier calipers after resection, recorded in submillimeters.
- **Paris classification**: According to the morphological characteristics under white light endoscopy, the European Society of Gastrointestinal Endoscopy (ESGE) recommends that the gross morphology of colorectal polyps should be described using the Paris classification system[46]. In our study, the Paris classification was assigned by the operating endoscopist based on the endoscopic appearance of each lesion. To ensure consistency and accuracy, a second experienced endoscopist was present throughout the procedure and independently verified each classification in real time.
- **Anatomical locations:** Determined in real time during colonoscopy using standard endoscopic cues (e.g., mucosal patterns, luminal landmarks, and centimeter markings on the colonoscope shaft to estimate the distance from the anal verge). Anatomical segmentation was based on the ten-segment division method proposed in the Endomapper dataset (ileum, ileocecal valve, cecum, ascending colon, hepatic flexure, transverse colon, splenic flexure, descending colon, sigmoid colon, and rectum)[47]. All labels were verified in real time by a second experienced endoscopist to ensure consistency.
- **Histological characteristics**: Determined through standardized pathological examination following formalin fixation. Polyps were classified into neoplastic and non-neoplastic categories according to established histopathological criteria[48].

    - Neoplastic polyps included two main types.

First, Adenomatous Polyps, which were further subtyped based on the proportion of villous architecture into Tubular Adenoma, Tubulovillous Adenoma, and Villous Adenoma[49].

Second, Serrated Lesions, which were classified according to the 2019 World Health Organization criteria as Hyperplastic Polyp, Sessile Serrated Lesion (SSL), Traditional Serrated Adenoma (TSA), and Serrated Adenoma-Unclassified (SAU)[48,50].

    - Non-neoplastic polyps included Inflammatory Polyps and Hamartomatous Polyps[48].

All histological labels were reviewed and validated by experienced gastrointestinal pathologists to ensure diagnostic accuracy and consistency.

For all critical assessment parameters (Paris classification, anatomical localization, and histopathological diagnosis), a dual-evaluator approach was implemented. Inter-observer agreement achieved 100% concordance across all three domains. Any initial discrepancies were resolved through expert consensus, ensuring robust data reliability and clinical validity.

Data Records

The Polyp-Size dataset is publicly accessible at Figshare[51] and consists of two primary components, as illustrated in the hierarchical structure in Fig. 4.

- Video recordings: Colonoscopy videos provided in a compressed archive titled Polyp_Size_Videos.zip. Each video is in MP4 format and follows a standardized naming convention (e.g., Video 01.mp4 through Video 42.mp4).
- Metadata file: A CSV file named Polyp_Size_Labels.csv containing detailed annotations for each video.

The metadata includes various fields organized by category (patient information, endoscopic equipment details, and clinical data), as detailed in Table 2. Each field in the metadata file has been carefully curated to ensure accuracy, consistency, and completeness. Table 3 provides representative examples from this metadata file, illustrating the diversity of polyp characteristics captured in the dataset. This comprehensive dataset supports the development and validation of AI-based systems for polyp size estimation and facilitates detailed analysis of polyp characteristics.

Technical Validation

Currently, there is no widely recognized method for polyp size regression. In polyp management, the 5 mm threshold serves as a crucial clinical marker: diminutive polyps ($\leq$5 mm) have minimal cancer risk and can potentially be resected without histopathological examination, while polyps larger than 5 mm require histological examination due to higher neoplastic potential[52–54]. This dichotomization guides key decisions like "resect-and-discard" and surveillance intervals[53]. Notably, the image features required for polyp size regression and classification tasks share significant similarities, and widely used binary classification models can be leveraged to validate the dataset's reliability. Given these considerations, we chose a polyp size classification task as a proxy for size regression to validate the dataset.

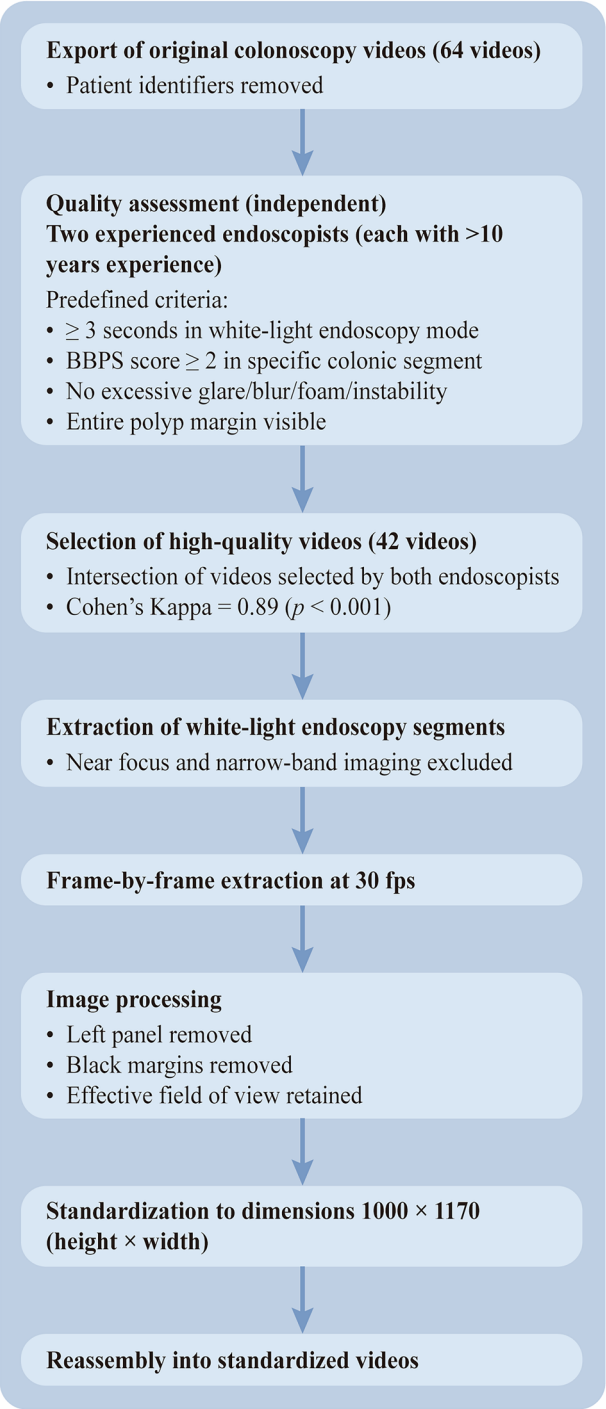

Fig. 2 Workflow of video preprocessing and standardization.

Firstly, we convert videos into image data at 15 fps and selected 3,858 images containing polyps to construct the training and validation datasets. The labels of the images indicate whether the size of the polyp in images is greater than 5 mm. Directly estimating the size of an object from a 2D image is an ill-posed problem. Therefore, we first obtain the metric depth maps of the images, which are fed into the classification network along with the processed images described above. Here we adopt the generalization metric depth estimation foundation model named ZoeN[55]. The predicted depth maps are visualized in Fig. 5. During training, the images and corresponding depth maps are concatenated to obtain four-dimensional RGBD input.

We separately employed the ResNet50[56], DenseNet169[57], and Inception V3[58] for binary classification tasks, which are widely used in previous polyp classification work[59–61] and are skilled in deep-level feature extraction. For accurate and reliable results, we carry out five-fold cross validation during training, the dataset split details are shown in Table 4.

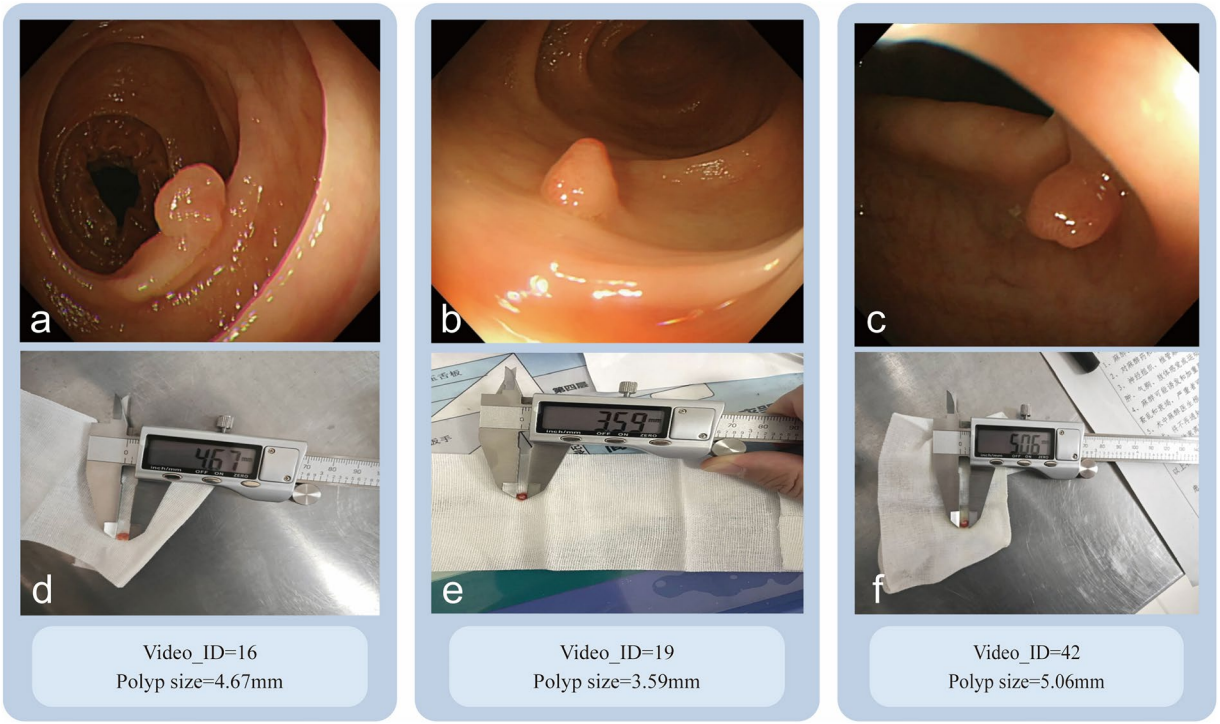

Fig. 3 Representative endoscopic images and corresponding physical measurements of colonic polyps from the Polyp-Size dataset. (**a**–**c**) Endoscopic views of three different polyps captured in the Polyp-Size video dataset. (**d**–**f**) Corresponding vernier caliper measurements of the resected specimens.

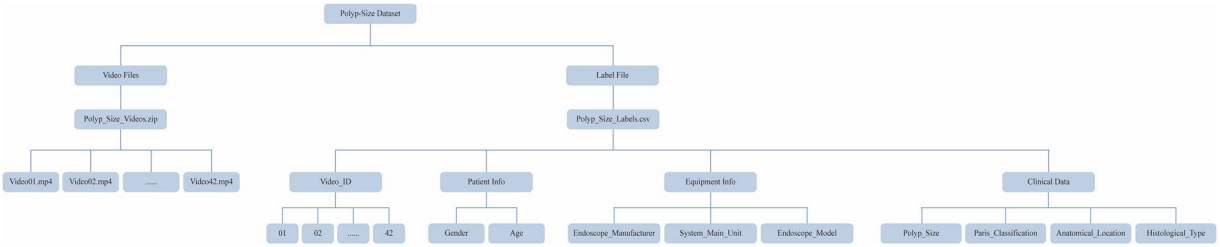

Fig. 4 Hierarchical structure of the Polyp-Size dataset, detailing its organization and components.

The experiences were conducted using the PyTorch (version 1.12.1) framework in Python 3.7.12 and accelerated using an NVIDIA GeForce RTX 4090 GPU. Before training, the model is initialized with the pre-trained weights from ImageNet[62]. The input layer of the model is modified to 4 channels for RGB-D input and initialized using Kaiming normal method. We conducted five-fold validation following Tian *et al.*[63]. For each fold, the network is trained for 50 epochs with cross-entropy loss and is optimized by the Adam optimizer, with an initial learning rate set to 0.001 according to previous work[64]. The batch size is set to 64 to balance the GPU memory limits and gradient stability. All activation functions used in the method maintain the original settings of the network models to avoid special designs and modifications. To facilitate better convergence, an ExponentialLR learning rate schedule is employed, where the learning rate is multiplied by 0.9 after each epoch. Additionally, before inputting data into the network, we apply a ColorJitter transformation to the RGB images with a probability of 0.5. We concatenate the RGB images (W*H*3) and depth maps (W*H*1) along the channel dimension to form a combined tensor (W*H*4), then apply random horizontal and vertical flips to ensure consistent and aligned transformations.

For evaluation metrics, classic binary classification metrics, including accuracy, F1 score, precision, recall, specificity and the area under ROC curve (AUC) are computed. The final result is the average value from five-fold cross-validation. Furthermore, we have also included the accuracy and loss plots for each epoch during training and validation as shown in Fig. 6, as well as ROC curves and detailed confusion matrices for each validation fold in Figs. 7, 8.

The results are shown in Table 5. All the different models can achieve an accuracy close to 0.65 when determining whether a polyp's size exceeds 5 mm and the performance differences among the models are relatively small. This result is generally acceptable, suggesting that the models can basically distinguishing between polyps

Category	Field Name	Description	Range/Values
Basic Information			
	Video_ID	A unique identifier linking each video to its corresponding metadata	01–42
	Gender	The gender of the patient	Male/Female
	Age	The patient's age at the time of the procedure	33–79 years
Equipment Information			
	Endoscope_Manufacturer	The manufacturer of the equipment	Olympus
	System_Main_Unit	The endoscope system processor model	CV-290
	Endoscope_Model	The specific colonoscope models used	CF-HQ290I, CF-H290I
Clinical Data			
	Polyp_Size	The diameter of the resected polyp, measured physically with a vernier caliper	2.80-11.74 mm
	Paris_Classification	The morphological classification of the polyp	Is, Ip, Isp, IIa
	Anatomical_Location	The polyp's location within the colon	Rectum, Sigmoid Colon, Descending Colon, Transverse Colon, etc
	Histological Type	The pathological classification	Tubular Adenoma, Sessile Serrated Lesion (SSL), Hyperplastic Polyp, etc.

Table 2. Comprehensive metadata fields in the Polyp-Size dataset including basic information, endoscopic equipment details, and clinical data.

of different sizes, which to some extent supports the reliability of the dataset. However, the current adopted metric depth estimation network we used still exhibits significant errors. Additionally, classic image classification networks such as ResNet50, DenseNet169, and Inception V3 are still insufficient in extracting image features. These factors have led to the current suboptimal results. We believe that the accuracy can be improved and the classification task can be extended to polyp size regression by designing more specialized metric depth estimation networks for endoscopic scenarios, incorporating more refined network architectures, and applying emerging techniques like 3D reconstruction.

Dataset Limitations

The Polyp-Size dataset provides a novel and practical resource for advancing colorectal polyp size measurement; however, it has some limitations. First, while the sample size is the largest among comparable datasets, it remains relatively modest and would benefit from further expansion to better support AI model training and validation. Second, data collection was limited to a single center, which may constrain the dataset's diversity and generalizability. This single-center limitation could impact the AI models' performance when applied to different populations or settings for several reasons: (1) The patient demographics at Renji Hospital may not fully represent global population diversity, with our cohort primarily consisting of East Asian individuals; (2) Institution-specific endoscopic techniques, equipment settings, and imaging protocols might influence image characteristics; and (3) Local clinical practices in polyp management could affect the spectrum of polyps included in the dataset. Our preliminary analysis of polyp size distribution across demographic variables revealed notable patterns. Among patients over 60 years, 31.3% presented with polyps >5 mm, compared to 42.3% in younger patients. Similarly, male patients had a higher proportion of polyps >5 mm (41.4%) than female patients (30.8%). While these differences did not reach statistical significance in our limited sample, they highlight potential demographic variations in polyp characteristics that a single-center dataset may not fully capture. These findings underscore the importance of developing multicenter datasets that include diverse populations to ensure AI models generalize effectively across different clinical settings and patient demographics. Additionally, in some cases, subjectivity in distinguishing polyp boundaries and subtle changes in polyp morphology or size due to retrieval through the endoscopic channel may introduce minor inaccuracies in size estimation. Besides, we adopted classification tasks using multiple widely-used models as a proxy for polyp size regression tasks to reliably demonstrate the dataset's usability in the Technical Validation section. While classification performance correlates with size estimation capability, it does not directly measure regression accuracy and provide doctors with accurate polyp size. We look forward to future researchers proposing solid polyp size regression algorithms based on our proposed dataset to further overcome the limitations of polyp size classification tasks, thereby providing physicians with accurate polyp size.

Despite these limitations, the Polyp-Size dataset provides a solid foundation for advancing intelligent polyp measurement technologies. Ongoing improvements in data collection and algorithmic refinement are expected to enhance its utility. We encourage researchers and clinical experts to utilize the dataset, offer constructive feedback, and contribute to its growth, supporting its continued development for colorectal cancer prevention and management.

In conclusion, the Polyp-Size dataset provides the first publicly available resource of high-precision polyp measurements, with key contributions including submillimeter-accurate vernier caliper validation, standardized endoscopic videos, comprehensive clinical metadata, and technical validation for size classification tasks. This resource addresses the critical challenge of subjective polyp sizing in clinical practice, potentially improving decision-making in surveillance planning and treatment selection. Future development will focus

Video_ID	Gender	Age	Endoscope_ Manufacturer	System_ Main_Unit	Endoscope_ Model	Polyp_ Size	Paris_ Classification	Anatomical_Location	Histological_Type
04	Male	59	Olympus	CV-290	CF-H290I	7.48 mm	IIa	Descending Colon	Tubular Adenoma
06	Male	56	Olympus	CV-290	CF-H290I	6.54 mm	IIa	Transverse Colon	Sessile Serrated Lesion (SSL)
19	Female	66	Olympus	CV-290	CF-HQ290I	3.59 mm	Is	Transverse Colon	Hyperplastic Polyp
29	Female	79	Olympus	CV-290	CF-HQ290I	9.31 mm	Is	Sigmoid Colon	Tubulovillous Adenoma
32	Male	61	Olympus	CV-290	CF-HQ290I	4.13 mm	Is	Sigmoid Colon	Inflammatory Polyp
41	Male	60	Olympus	CV-290	CF-HQ290I	5.17 mm	IIa	Ascending Colon	Tubular Adenoma

Table 3. Representative examples from the Polyp-Size dataset metadata (Polyp_Size_Labels.csv).

Fold	Number of patients in the training set	Number of images in the training set	Number of patients in the testing set	Number of images in the testing set
1	33	3190	9	668
2	33	2454	9	1404
3	34	3171	8	687
4	34	3245	8	613
5	34	3372	8	486

Table 4. The dataset split details: the number of patients and images in the training set and the test set during the five-fold cross-validation.

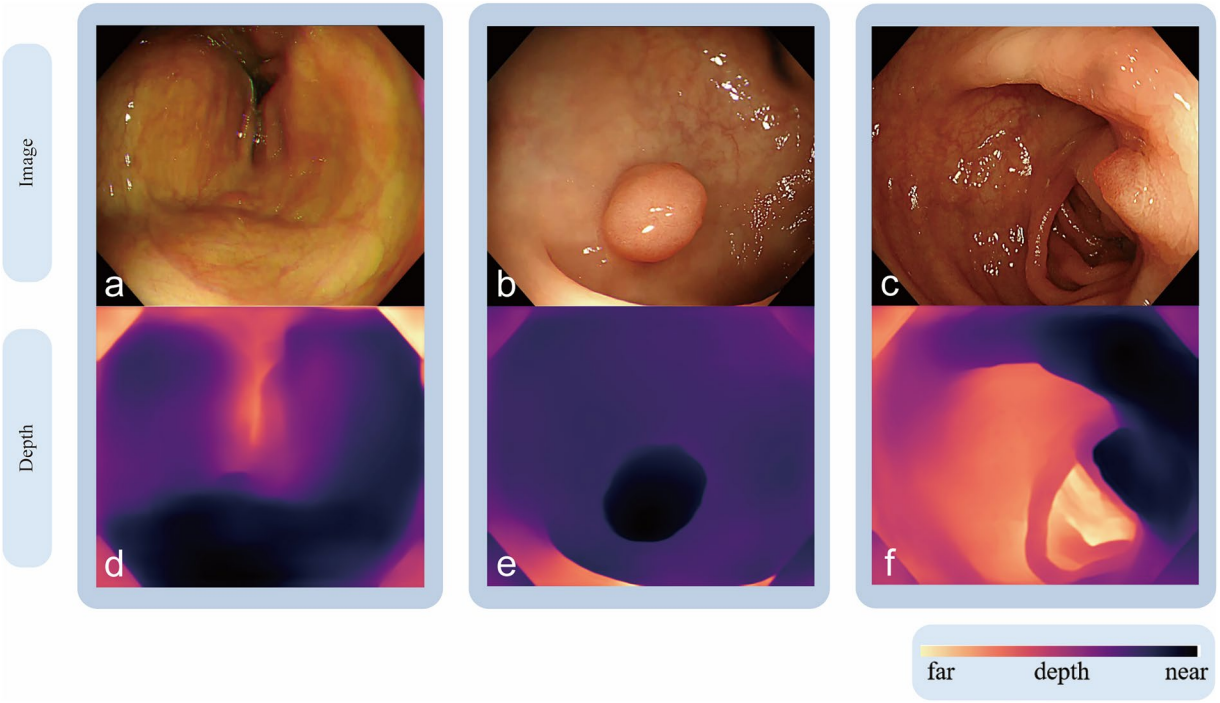

Fig. 5 Visualization of endoscopic images and their corresponding metric depth maps. (**a**–**c**) Representative endoscopic images of colonic polyps. (**d**–**f**) Depth maps generated by the ZoeN model, where brighter regions indicate greater distances from the camera. A scale bar represents relative depth progression from near (black) to far (white).

on multi-center data collection to enhance demographic and equipment diversity across different healthcare settings. We plan to integrate complementary imaging techniques such as narrow-band imaging (NBI) and magnifying endoscopy to improve visualization of polyp surface patterns. Additionally, we will develop specialized computational methods addressing endoscopic-specific challenges, including reflection handling and depth estimation algorithms optimized for the unique conditions of the gastrointestinal tract. Combining computer vision with clinical metadata could further enrich polyp characterization beyond size alone. Researchers interested in contributing to future iterations are encouraged to contact the corresponding authors with proposals for multi-center validation, algorithm development, or complementary data collection. Through this collaborative

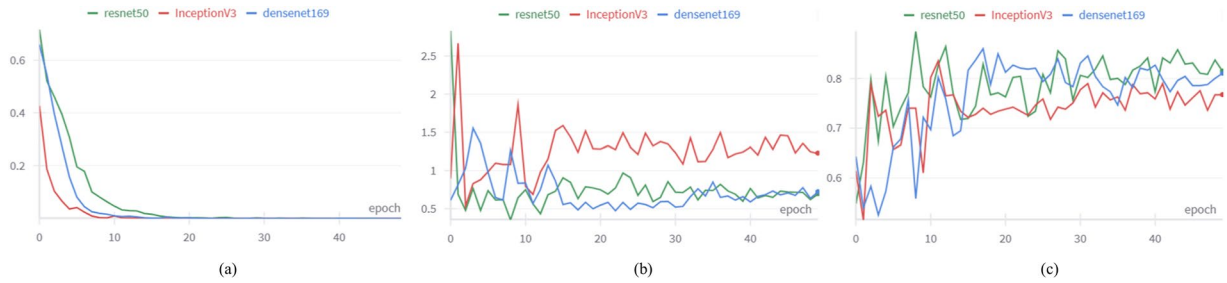

Fig. 6 The accuracy and loss plots for each epoch during training and validation for different models. (**a**) training loss plot. (**b**) validation loss plot. (**c**) validation accuracy plot.

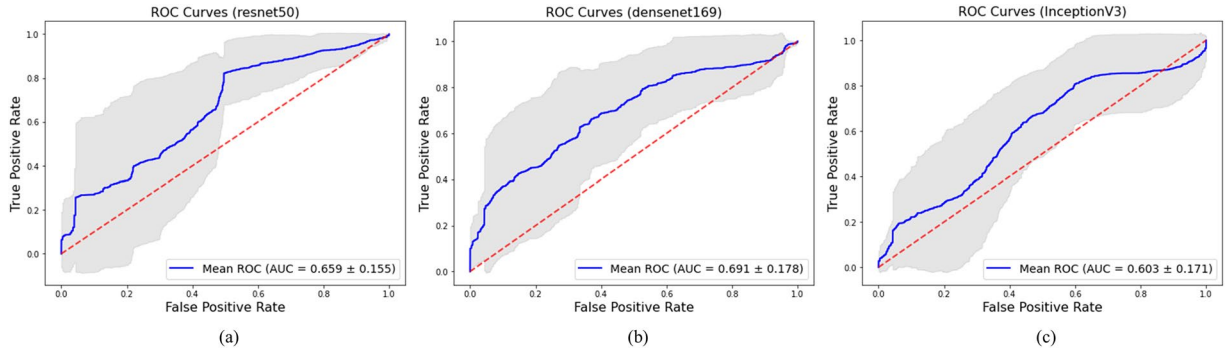

Fig. 7 The ROC curves for different models. (**a**) ROC curves for Resnet50. (**b**) ROC curves for Densenet169. (**c**) ROC curves for Inception3. The red diagonal line represents AUC = 0.5, the blue curve means the mean ROC across all validation folds and the shaded gray region shows the standard deviation range.

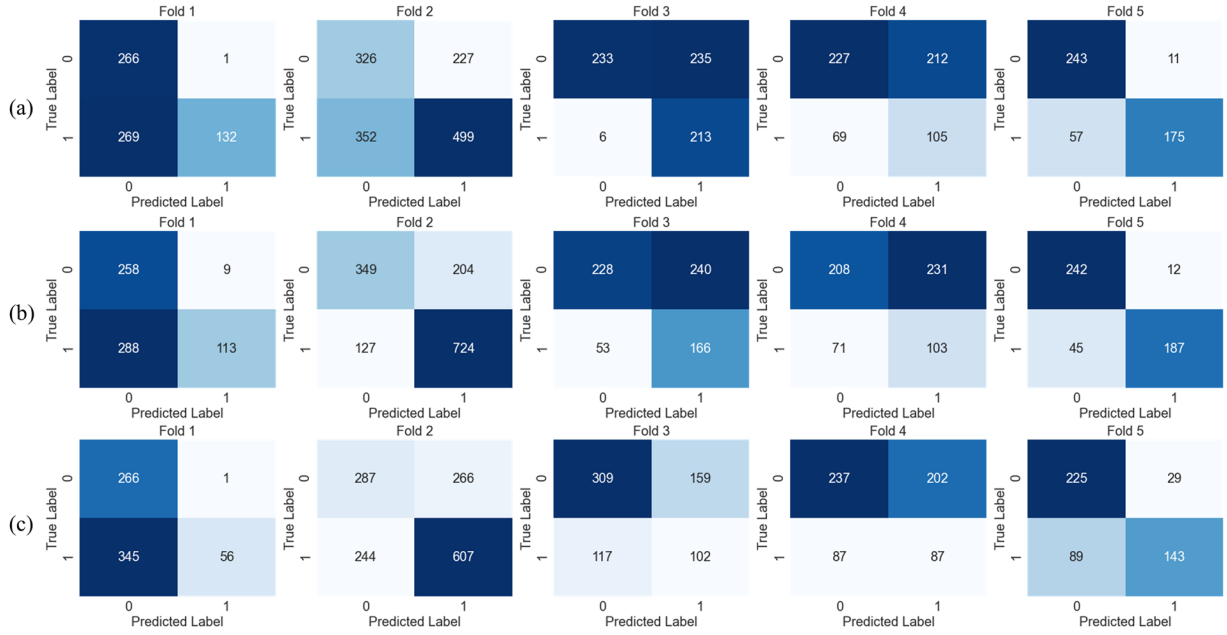

Fig. 8 The confusion matrices of five folds for different models Class 0 means the polyp sizes $\leq$ 5 mm and Class 1 means the polyp sizes exceed 5 mm. (**a**) confusion matrices for Resnet50. (**b**) confusion matrices for Densenet169. (**c**) confusion matrices for Inception3.

approach, the Polyp-Size dataset aims to advance precision medicine in colorectal cancer prevention by promoting standardized, AI-assisted measurement techniques globally.

Methods	Accuracy	Recall	Precision	F1 score	Specificity	AUC
Inception V3	0.601	0.487	0.640	0.491	0.720	0.603
DenseNet169	0.657	0.658	0.673	0.610	0.702	0.691
ResNet50	0.647	0.649	0.685	0.606	0.711	0.659

Table 5. Results of binary classification tasks for polyp size estimation using various methods, comparing accuracy, precision, recall, F1 scores, specificity and AUC.

Code availability

To facilitate dataset utilization, the preprocessing code is available at https://github.com/Scarlett213/Polyp-Size-Dataset. This code supports the conversion of video data into image datasets and splitting the data into training and validation sets. Detailed instructions for preprocessing and further customization are provided in the accompanying README.md file.

Received: 20 December 2024; Accepted: 21 May 2025;
Published online: 31 May 2025

References

1. Strum, W. B. Colorectal Adenomas. *N Engl J Med* **374**, 1065–1075, https://doi.org/10.1056/NEJMra1513581 (2016).
2. Mezzapesa, M. *et al*. Serrated Colorectal Lesions: An Up-to-Date Review from Histological Pattern to Molecular Pathogenesis. *Int J Mol Sci* **23**, https://doi.org/10.3390/ijms23084461 (2022).
3. Pan, J., Xin, L., Ma, Y. F., Hu, L. H. & Li, Z. S. Colonoscopy Reduces Colorectal Cancer Incidence and Mortality in Patients With Non-Malignant Findings: A Meta-Analysis. *Am J Gastroenterol* **111**, 355–365, https://doi.org/10.1038/ajg.2015.418 (2016).
4. Ladabaum, U., Dominitz, J. A., Kahi, C. & Schoen, R. E. Strategies for Colorectal Cancer Screening. *Gastroenterology* **158**, 418–432, https://doi.org/10.1053/j.gastro.2019.06.043 (2020).
5. Zhang, C., Cifu, A. S. & Patel, A. Recommendations for Follow-up Colonoscopy After Polypectomy. *Jama* **324**, 2208–2209, https://doi.org/10.1001/jama.2020.15001 (2020).
6. Copland, A. P., Kahi, C. J., Ko, C. W. & Ginsberg, G. G. AGA Clinical Practice Update on Appropriate and Tailored Polypectomy: Expert Review. *Clin Gastroenterol Hepatol* **22**, 470–479.e475, https://doi.org/10.1016/j.cgh.2023.10.012 (2024).
7. Ferlitsch, M. *et al*. Colorectal polypectomy and endoscopic mucosal resection (EMR): European Society of Gastrointestinal Endoscopy (ESGE) Clinical Guideline. *Endoscopy* **49**, 270–297, https://doi.org/10.1055/s-0043-102569 (2017).
8. Lieberman, D. A. *et al*. Guidelines for colonoscopy surveillance after screening and polypectomy: a consensus update by the US Multi-Society Task Force on Colorectal Cancer. *Gastroenterology* **143**, 844–857, https://doi.org/10.1053/j.gastro.2012.06.001 (2012).
9. Gupta, S. *et al*. Recommendations for Follow-Up After Colonoscopy and Polypectomy: A Consensus Update by the US Multi-Society Task Force on Colorectal Cancer. *Gastroenterology* **158**, 1131–1153.e1135, https://doi.org/10.1053/j.gastro.2019.10.026 (2020).
10. Tanaka, S. *et al*. Evidence-based clinical practice guidelines for management of colorectal polyps. *J Gastroenterol* **56**, 323–335, https://doi.org/10.1007/s00535-021-01776-1 (2021).
11. Chaptini, L. *et al*. Variation in polyp size estimation among endoscopists and impact on surveillance intervals. *Gastrointest Endosc* **80**, 652–659, https://doi.org/10.1016/j.gie.2014.01.053 (2014).
12. Cheloff, A. Z., Kim, L., Pochapin, M. B., Shaukat, A. & Popov, V. Accuracy of Visual Estimation for Measuring Colonic Polyp Size: A Systematic Review and Meta-Analysis. *Am J Gastroenterol* https://doi.org/10.14309/ajg.0000000000003391 (2025).
13. Summers, R. M. Polyp size measurement at CT colonography: what do we know and what do we need to know? *Radiology* **255**, 707–720, https://doi.org/10.1148/radiol.10090877 (2010).
14. Wang, J. *et al*. A real-time deep learning-based system for colorectal polyp size estimation by white-light endoscopy: development and multicenter prospective validation. *Endoscopy* **56**, 260–270, https://doi.org/10.1055/a-2189-7036 (2024).
15. Krenzer, A., Hekalo, A. & Puppe, F. Endoscopic Detection And Segmentation Of Gastroenterological Diseases With Deep Convolutional Neural Networks. In *EndoCV@ ISBI*, 58–63 (2020).
16. Krenzer, A. *et al*. Automated classification of polyps using deep learning architectures and few-shot learning. *BMC Medical Imaging* **23**, 59, https://doi.org/10.1186/s12880-023-01007-4 (2023).
17. Tudela, Y. *et al*. A complete benchmark for polyp detection, segmentation and classification in colonoscopy images. *Frontiers in oncology* **14**, 1417862, https://doi.org/10.3389/fonc.2024.1417862 (2024).
18. Livovsky, D. M. *et al*. Detection of elusive polyps using a large-scale artificial intelligence system (with videos). *Gastrointest Endosc* **94**, 1099–1109.e1010, https://doi.org/10.1016/j.gie.2021.06.021 (2021).
19. Hassan, C. *et al*. Performance of artificial intelligence in colonoscopy for adenoma and polyp detection: a systematic review and meta-analysis. *Gastrointest Endosc* **93**, 77–85.e76, https://doi.org/10.1016/j.gie.2020.06.059 (2021).
20. Krenzer, A. *et al*. A Real-Time Polyp-Detection System with Clinical Application in Colonoscopy Using Deep Convolutional Neural Networks. *J Imaging* **9**, https://doi.org/10.3390/jimaging9020026 (2023).
21. Zhang, X. *et al*. Real-time gastric polyp detection using convolutional neural networks. *PLoS One* **14**, e0214133, https://doi.org/10.1371/journal.pone.0214133 (2019).
22. Misawa, M. *et al*. Artificial Intelligence-Assisted Polyp Detection for Colonoscopy: Initial Experience. *Gastroenterology* **154**, 2027–2029.e2023, https://doi.org/10.1053/j.gastro.2018.04.003 (2018).
23. Wan, J., Chen, B. & Yu, Y. Polyp Detection from Colorectum Images by Using Attentive YOLOv5. *Diagnostics (Basel)* **11**, https://doi.org/10.3390/diagnostics11122264 (2021).
24. Troya, J. *et al*. Direct comparison of multiple computer-aided polyp detection systems. *Endoscopy* **56**, 63–69, https://doi.org/10.1055/a-2147-0571 (2024).
25. Brand, M. *et al*. Development and evaluation of a deep learning model to improve the usability of polyp detection systems during interventions. *United European Gastroenterol J* **10**, 477–484, https://doi.org/10.1002/ueg2.12235 (2022).
26. Brand, M. *et al*. Frame-by-Frame Analysis of a Commercially Available Artificial Intelligence Polyp Detection System in Full-Length Colonoscopies. *Digestion* **103**, 378–385, https://doi.org/10.1159/000525345 (2022).
27. Guo, X. *et al*. Automated polyp segmentation for colonoscopy images: A method based on convolutional neural networks and ensemble learning. *Med Phys* **46**, 5666–5676, https://doi.org/10.1002/mp.13865 (2019).
28. Ren, X. *et al*. Prompt-based polyp segmentation during endoscopy. *Med Image Anal* **102**, 103510, https://doi.org/10.1016/j.media.2025.103510 (2025).

29. Sudarevic, B. *et al*. Artificial intelligence-based polyp size measurement in gastrointestinal endoscopy using the auxiliary waterjet as a reference. *Endoscopy* **55**, 871–876, https://doi.org/10.1055/a-2077-7398 (2023).
30. 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, https://doi.org/10.1111/den.14318 (2022).
31. Eisenmann, M. *et al*. Biomedical image analysis competitions: The state of current participation practice. *arXiv preprint arXiv:2212.08568*, https://doi.org/10.48550/arXiv.2212.08568 (2022).
32. Krenzer, A. *et al*. Fast machine learning annotation in the medical domain: a semi-automated video annotation tool for gastroenterologists. *Biomed Eng Online* **21**, 33, https://doi.org/10.1186/s12938-022-01001-x (2022).
33. Tajbakhsh, N., Gurudu, S. R. & Liang, J. Automated Polyp Detection in Colonoscopy Videos Using Shape and Context Information. *IEEE Trans Med Imaging* **35**, 630–644, https://doi.org/10.1109/tmi.2015.2487997 (2016).
34. Bernal, J. *et al*. WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. *Comput Med Imaging Graph* **43**, 99–111, https://doi.org/10.1016/j.compmedimag.2015.02.007 (2015).
35. Vázquez, D. *et al*. A benchmark for endoluminal scene segmentation of colonoscopy images. *Journal of healthcare engineering* **2017**, 4037190, https://doi.org/10.1155/2017/4037190 (2017).
36. Jha, D. *et al*. Kvasir-seg: A segmented polyp dataset. In *MultiMedia modeling: 26th international conference, MMM 2020*, 451–462 (2020).
37. Sánchez-Peralta, L. F. *et al*. PICCOLO White-Light and Narrow-Band Imaging Colonoscopic Dataset: A Performance Comparative of Models and Datasets. *Applied Sciences* **10**, 8501, https://doi.org/10.3390/app10238501 (2020).
38. Ali, S. *et al*. A multi-centre polyp detection and segmentation dataset for generalisability assessment. *Scientific Data* **10**, 75, https://doi.org/10.1038/s41597-023-01981-y (2023).
39. Wang, W. *et al*. An improved deep learning approach and its applications on colonic polyp images detection. *BMC Medical Imaging* **20**, 1–14, https://doi.org/10.1186/s12880-020-00482-3 (2020).
40. Mangotra, H., Handa, P. & Gooel, N. Open-Source Datasets for Colonoscopy Polyps and Its AI-Enabled Techniques. In *International Conference on Communication and Intelligent Systems*, 63–76 (2022).
41. Misawa, M. *et al*. Development of a computer-aided detection system for colonoscopy and a publicly accessible large colonoscopy video database (with video). *Gastrointestinal endoscopy* **93**, 960–967. e963, https://doi.org/10.1016/j.gie.2020.07.060 (2021).
42. Ma, Y., Chen, X., Cheng, K., Li, Y. & Sun, B. LDPolypVideo benchmark: a large-scale colonoscopy video dataset of diverse polyps. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference*, 387–396 (2021).
43. Fitting, D. *et al*. A video based benchmark data set (ENDOTEST) to evaluate computer-aided polyp detection systems. *Scandinavian Journal of Gastroenterology* **57**, 1397–1403, https://doi.org/10.1080/00365521.2022.2085059 (2022).
44. Li, K. *et al*. Colonoscopy polyp detection and classification: Dataset creation and comparative evaluations. *Plos one* **16**, e0255809, https://doi.org/10.1371/journal.pone.0255809 (2021).
45. Pacal, I. & Karaboga, D. A robust real-time deep learning based automatic polyp detection system. *Comput Biol Med* **134**, 104519, https://doi.org/10.1016/j.compbiomed.2021.104519 (2021).
46. Update on the paris classification of superficial neoplastic lesions in the digestive tract. *Endoscopy* **37**, 570-578, https://doi.org/10.1055/s-2005-861352 (2005).
47. Azagra, P. *et al*. Endomapper dataset of complete calibrated endoscopy procedures. *Sci Data* **10**, 671, https://doi.org/10.1038/s41597-023-02564-7 (2023).
48. Wang, R. J. *et al*. Standardized diagnosis and treatment of colorectal polyps. *Zhonghua Wei Chang Wai Ke Za Zhi* **27**, 583–590, https://doi.org/10.3760/cma.j.cn441530-20240416-00143 (2024).
49. Ishii, T. *et al*. Tubular adenomas with minor villous changes show molecular features characteristic of tubulovillous adenomas. *Am J Surg Pathol* **35**, 212–220, https://doi.org/10.1097/PAS.0b013e318205df20 (2011).
50. Ahadi, M., Sokolova, A., Brown, I., Chou, A. & Gill, A. J. The 2019 World Health Organization Classification of appendiceal, colorectal and anal canal tumours: an update and critical assessment. *Pathology* **53**, 454–461, https://doi.org/10.1016/j.pathol.2020.10.010 (2021).
51. Song, Y.-M. *et al*. Polyp-Size: A Precise Endoscopic Dataset for AI-Driven Polyp Sizing. *figshare* https://doi.org/10.6084/m9.figshare.28030115 (2025).
52. Sneh Arbib, O. *et al*. Risk of advanced lesions at the first follow-up colonoscopy after polypectomy of diminutive versus small adenomatous polyps of low-grade dysplasia. *Gastrointest Endosc* **86**, 713–721.e712, https://doi.org/10.1016/j.gie.2017.02.034 (2017).
53. Abu Dayyeh, B. K. *et al*. ASGE Technology Committee systematic review and meta-analysis assessing the ASGE PIVI thresholds for adopting real-time endoscopic assessment of the histology of diminutive colorectal polyps. *Gastrointest Endosc* **81**, 502.e501–502.e516, https://doi.org/10.1016/j.gie.2014.12.022 (2015).
54. Ponugoti, P. L., Cummings, O. W. & Rex, D. K. Risk of cancer in small and diminutive colorectal polyps. *Dig Liver Dis* **49**, 34–37, https://doi.org/10.1016/j.dld.2016.06.025 (2017).
55. Bhat, S. F., Birkl, R., Wofk, D., Wonka, P. & Müller, M. Zoedepth: Zero-shot transfer by combining relative and metric depth. *arXiv preprint arXiv:2302.12288*, https://doi.org/10.48550/arXiv.2302.12288 (2023).
56. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778 (2016).
57. Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K. Q. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4700–4708 (2017).
58. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2818–2826 (2016).
59. Bour, A., Castillo-Olea, C., Garcia-Zapirain, B. & Zahia, S. Automatic colon polyp classification using convolutional neural network: a case study at basque country. In *2019 IEEE international symposium on signal processing and information technology (ISSPIT)*, 1–5 (2019).
60. Patel, K. *et al*. A comparative study on polyp classification using convolutional neural networks. *PloS one* **15**, e0236452, https://doi.org/10.1371/journal.pone.0236452 (2020).
61. Lo, C.-M., Yeh, Y.-H., Tang, J.-H., Chang, C.-C. & Yeh, H.-J. Rapid polyp classification in colonoscopy using textural and convolutional features. In *Healthcare*, **1494** (2022).
62. Deng, J. *et al*. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, 248–255 (2009).
63. Tian, Y., Pu, L. Z., Singh, R., Burt, A. D. & Carneiro, G. One-stage five-class polyp detection and classification. In *2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019)*, 70–73 (2019).
64. Hossain, M. B., Iqbal, S. H. S., Islam, M. M., Akhtar, M. N. & Sarker, I. H. Transfer learning with fine-tuned deep CNN ResNet50 model for classifying COVID-19 from chest X-ray images. *Informatics in Medicine Unlocked* **30**, 100916, https://doi.org/10.1016/j.imu.2022.100916 (2022).
65. Forootan, M. *et al*. ERCPMP: an endoscopic image and video dataset for colorectal polyps morphology and pathology. *BMC Res Notes* **17**, 393, https://doi.org/10.1186/s13104-024-07062-6 (2024).
66. Biffi, C. *et al*. REAL-Colon: A dataset for developing real-world AI applications in colonoscopy. *Sci Data* **11**, 539, https://doi.org/10.1038/s41597-024-03359-0 (2024).

Acknowledgements
This work was financially supported by Science and Technology Commission of Shanghai Municipality (Grant No. 23Y11905300).

Author contributions
Yiming Song initiated and coordinated the work on the Polyp-Size dataset, collected colonoscopy data, curated video recordings, organized the dataset, and wrote the manuscript. Sijia Du conducted the video standardization process, performed technical validation, and wrote the manuscript. Ruilan Wang and Fei Liu established the standardized protocol for polyp size measurement and performed metadata quality control. Xiaolu Lin and Jinnan Chen collated and organized information from previous polyp datasets and created corresponding tables. Zeyu Li and Zhao Li were responsible for collecting the colonoscopy videos and coordinating metadata collection. Liuyi Yang collected the colonoscopy videos and verified the endoscopic equipment specifications. Zhengjie Zhang analyzed the initial data and coordinated the standardization process. Hao Yan was responsible for the design and drawing of the figures. Qingwei Zhang organized endoscopic experts to select the collected videos, developed the criteria for video selection, and performed video quality control. Dahong Qian supervised the technical validation methodology and guided the standardization process of the dataset. Xiaobo Li performed the colonoscopy procedures, organized the completion of polyp size measurements, and organized the overall research implementation.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to Q.Z., D.Q. or X.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.