

Development and Validation of SERRATE: A Deep Learning-Based Tool for Diagnosis of Sessile Serrated Lesions in Colonoscopy Images

Wenmo Hu

Peking Union Medical College Hospital

Erdong Chen

University of Massachusetts Lowell

Linfeng Zou

Peking Union Medical College Hospital

Qilei Chen

University of Massachusetts Lowell

Tianrui Hua

Peking Union Medical College Hospital

Kai Song

Peking Union Medical College Hospital

Yanyan Wu

Peking Union Medical College Hospital

Long Zou

Peking Union Medical College Hospital

Pengguang Yan

Peking Union Medical College Hospital

Bowen Tian

Peking Union Medical College Hospital

Mengmeng Zhang

Peking Union Medical College Hospital

Zhaohui Fang

Ningcheng County Central Hospital

Na Yang

Jinan Third People's Hospital

Xiaoyan Yang

Chengdu Second People's Hospital

Yafeng Lu

Qingyang Municipal People's Hospital

Minglan Yang

Linyi Central hospital

Yan Zhao

Zhengzhou Seventh People's Hospital

Jianing Li

Peking Union Medical College Hospital

Shengyu Zhang

Peking Union Medical College Hospital

Qingwei Jiang

Peking Union Medical College Hospital

Benyuan Liu

University of Massachusetts Lowell

Yu Cao

University of Massachusetts Lowell

Yingyun Yang

Peking Union Medical College Hospital

Dong Wu

wudong061002@126.com

Peking Union Medical College Hospital

Aiming Yang

Peking Union Medical College Hospital

Research Article

Keywords: Sessile serrated lesions, Colonoscopy, Polyp Classification, Computer-Aided Diagnosis, Deep Learning

Posted Date: August 13th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10122585/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Development and Validation of SERRATE: A Deep Learning-Based Tool for Diagnosis of Sessile Serrated Lesions in Colonoscopy Images

Wenmo Hu^{1,#}, Erdong Chen^{4,#}, Linfeng Zou^{1,#}, Qilei Chen⁴, Tianrui Hua¹, Kai Song¹, Yanyan Wu², Long Zou⁵, Pengguang Yan², Bowen Tian¹, Mengmeng Zhang², Zhaohui Fang⁶, Na Yang⁷, Xiaoyan Yang⁸, Yafeng Lu⁹, Minglan Yang¹⁰, Yan Zhao¹¹, Jianing Li², Shengyu Zhang², Qingwei Jiang², Benyuan Liu⁴, Yu Cao⁴, Yingyun Yang^{2,*}, Dong Wu^{2,3,*}, Aiming Yang²

1. Department of Internal Medicine, State Key Laboratory of Complex Severe and Rare Diseases, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Peking, China.
2. Department of Gastroenterology, State Key Laboratory of Complex Severe and Rare Diseases, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Peking, China.
3. Clinical Epidemiology Unit, International Clinical Epidemiology Network, Beijing, China.
4. Department of Computer Science, University of Massachusetts Lowell, Lowell, MA, USA.
5. Department of Pathology, State Key Laboratory of Complex Severe and Rare Diseases, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Peking, China.
6. Department of Gastroenterology, Ningcheng County Central Hospital, Chifeng City, Inner Mongolia Autonomous Region, China.
7. Department of Gastroenterology, Jinan Third People's Hospital, Jinan, Shandong, China.
8. Department of Gastroenterology, West China School of Medicine, Sichuan University, Sichuan University affiliated Chengdu Second People's Hospital, Chengdu Second People's Hospital, Chengdu, China.
9. Department of Gastroenterology, Qingyang Municipal People's Hospital, Qingyang, Gansu, China.

10. Department of Gastroenterology, Linyi Central hospital, Linyi, Shandong, China.
11. Department of Gastroenterology, Zhengzhou Seventh People's Hospital, Zhengzhou, Henan, China.

[#]These authors contributed equally: Wenmo Hu, Erdong Chen, Linfeng Zou

*Correspondence authors

No.1 Shuaifuyuan, Dongcheng District, Beijing 100730, China.

Email: yangyingyun@pumch.cn; yang.yingyun@mayo.edu; wudong@pumch.cn

Tel: 0086-18612671010

Fax: 0086-10-69155017

Funding

(1) Excellence Clinical Research Program for Research-based Wards (No. BRWEP2024W034010103).

(2) National Key Research and Development Program of China (No. 2024YFA0918504).

(3) Chinese Academy of Medical Sciences (CAMS) Innovation Fund for Medical Sciences (No. 2022-I2M-1-003).

(4) National High Level Hospital Clinical Research Funding (2022-PUMCH-B-024, 2022-PUMCH-A-071).

(5) Peking Union Medical College Hospital Talent Cultivation Program (Category C) (No.UBJ11356).

Abstract

Background and Aims: Sessile serrated lesions (SSLs) are precancerous colorectal polyps with high malignant potential, yet their endoscopic recognition remains challenging. Most artificial intelligence (AI) models for polyp classification fail to specifically identify SSLs, often misclassifying them as hyperplastic polyps. This study aimed to develop and validate SERRATE (Sessile Serrated lesion Recognition and Risk Assessment Tool for Endoscopy), a deep learning model for distinguishing SSLs from conventional adenomas and hyperplastic polyps.

Methods: We retrospectively collected colonoscopic images from 696T patients at Peking Union Medical College Hospital (January 2017-September 2024), yielding >20,000 images. The balanced training dataset comprised 1,120 SSL images, 1,697 conventional adenoma images, and 1,207 hyperplastic polyp images. A Swin Transformer model was trained using both white-light imaging and narrow-band imaging (NBI). Model performance was evaluated on a prospective internal validation cohort (300 patients, January 2025-April 2025; ClinicalTrials.gov NCT06773832) and the external POLAR dataset.

Results: The optimal model utilized NBI images with $2\times$ mucosal context expansion, achieving 82.42% accuracy in internal validation. For three-class classification (SSL vs. conventional adenoma vs. hyperplastic polyp), the model demonstrated 73.44%

overall accuracy (precision 82.47%, recall 95.21%, specificity 91.21%, F1-score 82.17%) on prospective internal validation and 69.46% accuracy on external validation. For binary neoplasia detection, accuracy reached 91.07% across validation sets.

Conclusions: SERRATE represents the first AI model specifically designed for SSL recognition with balanced training data. The model achieved clinically relevant performance for SSL differentiation, potentially improving colorectal cancer prevention through enhanced serrated lesion management.

Keywords: Sessile serrated lesions, Colonoscopy, Polyp Classification, Computer-Aided Diagnosis, Deep Learning

1. Introduction

Colorectal cancer (CRC) is a major global health burden, ranking third in incidence and second in cancer-related mortality, with 1.9 million new cases and 0.9 million deaths reported in 2020¹. Alarmingly, these numbers are predicted to rise to 3.2 million cases and 1.6 million deaths by 2040². This escalating burden underscores the critical importance of effective screening and early detection strategies, particularly through colonoscopy-based surveillance programs.

Although the conventional adenoma pathway accounts for the majority of colorectal cancer (CRC) cases, the serrated pathway—which contributes to approximately 15–30% of CRCs—also warrants significant attention due to its unique diagnostic challenges. Serrated lesions including sessile serrated lesion (SSL) and traditional serrated adenomas (TSA) are considered precursor lesions to CRC and progress more rapidly toward malignancy³. Besides, SSL is considerably more difficult to detect than conventional adenomas due to its flat morphology, subtle endoscopic appearance, and frequent location in the proximal colon⁴.

Moreover, SSLs are often difficult to distinguish morphologically from hyperplastic polyps, which are benign and non-neoplastic. Even among experienced gastrointestinal pathologists, diagnostic concordance based on histopathological evaluation remains limited⁵. This places greater demands on endoscopists to

accurately identify and appropriately manage lesions during colonoscopy—whereas small hyperplastic polyps may be left untreated, SSLs require complete resection.

Universal resection of all detected lesions may reduce the risk of missing SSLs, but at the cost of increased burden to patients and healthcare systems. Conversely, failure to remove SSLs poses a substantial risk of malignant transformation. Therefore, improving the detection and management of SSLs is essential for effective CRC prevention.

The application of artificial intelligence (AI) offers promising potential to address the current limitations in endoscopic recognition of serrated lesions. In recent years, computer-aided diagnosis (CADx) systems have advanced rapidly. Multiple large-scale, prospective studies have demonstrated that CADx achieves diagnostic accuracy for optical characterization of diminutive lesions (<5 mm) that meets or exceeds the performance thresholds set by major European and American endoscopy societies^{6–11}.

However, most current research focuses primarily on differentiating conventional adenomas from hyperplastic polyps, with limited attention given to serrated lesions, particularly sessile serrated lesions (SSLs). Notably, some commercially available CADx systems misclassify most SSLs as hyperplastic polyps, potentially leading endoscopists to make inappropriate management decisions and increasing the risk of

neoplastic progression^{8,12}.

One major reason for this limitation is the inadequate representation of SSLs in training datasets. SSLs may have been excluded altogether, incorrectly labeled, or included in insufficient numbers to meet the data volume requirements of AI-based clustering and classification tasks. As a result, current models lack the necessary exposure and learning capacity to reliably identify SSLs in clinical practice^{10,11}.

We therefore developed SERRATE () at PUMCH, a CADx model targeting SSL differentiation with a class-balanced dataset including conventional adenomas and hyperplastic polyps, evaluated on prospective internal and external validation cohorts.

2. Patients and methods

2.1 Study design

2.1.1 Phase 1: Single-center Retrospective AI Model Development

To focus on improving the diagnostic accuracy for serrated lesions, we retrospectively collected cases diagnosed up to September 2024 at PUMCH. These included serrated lesions including SSAs and TSAs, all confirmed by gastrointestinal pathology experts with independent review of each diagnosis.

To optimize the discriminative capability of the AI model, we constructed a balanced dataset by matching each serrated lesion with a comparable number of hyperplastic polyps and conventional adenomas. Each lesion was represented by one or more high-resolution white-light images, with or without narrow-band imaging (NBI), acquired using the Olympus CV-290 system.

Two experienced endoscopists (each with >10 years of experience and >2,000 colonoscopies performed) annotated all images. Annotations were based on the colonoscopy reports and corresponding pathology reports to determine the definitive pathological diagnosis for each image. Images were excluded if the pathological diagnosis could not be clearly established from the available documentation. In cases where the two endoscopists provided conflicting annotations, a third expert endoscopist was consulted. Only images for which all three experts reached a consensus were retained; otherwise, the images were excluded from further analysis.

Within each lesion category, images were selected consecutively from all eligible cases identified during the study period (January 2017–September 2024), in chronological order of colonoscopy date. Images were excluded only on the basis of predefined quality criteria: motion blur, inadequate lesion visualization, or inability to clearly establish pathological diagnosis from available documentation. No images were excluded on the basis of lesion appearance or to favor any particular

morphological pattern.

For image preprocessing, we employed a pretrained YOLOv8 (You Only Look Once) model to identify and crop the image to get effective content¹³. Then, a pretrained Mask Region-based Convolutional Neural Network (R-CNN) model was utilized to locate the lesions with the bounding boxes with the minimum region of surrounding mucosa ($1\times$)¹⁴. In addition, we also generated regions with 1.5 times ($1.5\times$) and 2 times ($2\times$) of surrounding mucosa to identify the best localization strategy.

We developed a CADx model to classify hyperplastic polyps, conventional adenomas, and SSLs, testing multiple deep learning architectures: ResNeXt50, ResNet50, EfficientNet, MobileNet, Vision Transformer (ViT), ConvNeXt, and Swin Transformer (Figure 1).

The annotated image dataset was partitioned at the patient level into a training subset (70%) and an internal validation subset (30%), with all images from the same lesion assigned exclusively to one subset to prevent data leakage. This internal validation subset was used solely for comparing model architectures and ROI magnification

strategies during model selection; it was not used for hyperparameter optimization.

The prospective internal validation cohort (Phase 2) and the external POLAR dataset were kept entirely independent from all training and model selection procedures and were accessed only after the final model configuration was determined.

The best-performing model on internal validation was selected for further evaluation in Phase 2.

2.1.2 Phase 2: Prospective internal validation and public dataset external validation

This prospective cohort study was registered at ClinicalTrials.gov (NCT06773832) in January 2025; enrollment concluded in April 2025.

Starting in January 2025, colonoscopy images were prospectively collected from patients undergoing endoscopic examinations at the Department of Gastroenterology of PUMCH. Patients undergoing colonoscopy for various clinical indications, were eligible for inclusion. By April 2025, we had included all images corresponding to histologically confirmed SSLs. For each SSL, we matched a comparable number of conventional adenomas and hyperplastic polyps diagnosed during the same time. All included lesions were completely resected and their pathological diagnoses were reviewed by gastrointestinal pathologists with subspecialty expertise.

For external validation, we tested the model using a publicly available dataset from the Netherlands—the POLAR dataset—originally constructed for the clinical validation of a real-time CADx system. This multicenter dataset consisted of 2,637 non-magnified narrow-band imaging (NBI) images from 1,339 histologically confirmed diminutive colorectal polyps collected across eight Dutch hospitals. The dataset included 73% conventional adenomas, 10% SSLs, and 17% hyperplastic polyps, and was designed to support development and benchmarking of CADx systems. Notably, all images in this dataset were captured using NBI, and the ground truth diagnoses were confirmed by expert gastrointestinal pathologists following standardized protocols¹⁵.

2.2 Outcome measurement

2.2.1 Primary Outcome

The primary outcome of the study is the diagnostic accuracy of the CADx model, with pathology as the gold standard.

2.2.2 Secondary Outcomes

Precision, recall, F1-score, and specificity were calculated for each validation dataset.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

2.3 Sample size calculation

The sample size calculation is performed using the two-tailed Chi-square test. The alternative hypothesis is that the diagnostic accuracy of the AI model differs from that of endoscopists in the pathological diagnosis of polyps. Based on previous study results, the accuracy rate of endoscopists in optical diagnosis (adenoma vs. non-adenoma) is around 70%^{9,16}. We anticipate that the accuracy of endoscopists in further distinguishing serrated adenomas with conventional adenomas and hyperplastic polyps to be around 60%. Our study aims to observe an improvement in diagnostic accuracy by AI model of 20% compared to endoscopists. To achieve the desired statistical power of 80% and significance level of 0.05, a total of 91 polyp lesions need to be prospectively included in the study. With a median adenoma detection rate of approximately 20%^{17,18}, a total of 455 colonoscopies are anticipated to be required.

2.4 Statistical analyses

Continuous data were shown as the mean $\pm$ standard deviation or median with interquartile range (IQR)], whereas categorical variables were reported as percentages. For the optical diagnosis judgments made by the CADx model, the diagnostic accuracy, precision, recall, F1-score, and specificity were calculated to reflect diagnostic performance. All tests with a two-sided $p < 0.05$ were considered statistically significant. Statistical analyses were conducted with SPSS version 25.

Patient and Public Involvement

Patients or the public were not involved in the design, conduct, reporting, or dissemination plans of this research.

3. Results

3.1 Baseline Characteristics of Patients and Datasets

Colonoscopic images were retrospectively collected from 696 patients at PUMCH (January 2017–September 2024), yielding more than 20,000 images. Ultimately, the training dataset consisted of 1,207 images of hyperplastic polyps (664 white-light, 543 NBI), 1,697 images of conventional adenomas (1,409 white-light, 288 NBI), and 1,120 SSL images (706 white-light, 414 NBI). The detailed distribution of images is

presented in Table 1.

For prospective internal validation, colonoscopic images from an additional 300 patients at PUMCH were collected between January 2025 and April 2025. Among these, 42 patients had SSL lesions confirmed by pathology after complete excision. The distribution of this prospective internal validation dataset is detailed in Table 2. External validation was conducted using the publicly available POLAR dataset published by Houwen *et al* in 2023¹⁵.

3.2 Study Outcomes

3.2.1 Performance of Different AI Models on Internal Dataset

All models were trained on preprocessed endoscopic images.

As summarized in Table 3, the models were compared in terms of accuracy, precision, recall (sensitivity), specificity, and F1-score on the internal validation set. Among the architectures tested, the Swin Transformer–based model demonstrated superior performance across all evaluated metrics, with a corresponding accuracy of 73.19%, precision of 71.22%, recall of 70.51%, specificity of 85.25%, and F1-score of 70.08%. The Swin Transformer was selected for model refining and prospective internal validation and external testing.

We then refined the Swin model by feeding with the lesions containing varying amounts of surrounding mucosa (crop scales of $1\times$, $1.5\times$, and $2\times$), to assess the influence of contextual information on classification performance. The results were summarized in Table 4. Notably, training with input images that included expanded mucosal context ($1.5\times$ or $2\times$ crop) yielded higher classification accuracy than that with limited mucosal context ($1\times$ crop). Overall, the model with NBI image and $2\times$ magnification performed the best with accuracy of 82.42%, precision of 82.47%, recall of 95.21%, specificity of 91.21%, and F1-score of 82.17% (Figure 2).

Gradient-based heatmap visualization (Figure 3 presents the outputs of the Swin model outputs for representative polyp images. Heatmaps were predominantly concentrated on lesion regions, supporting model interpretability, while activation in surrounding mucosa confirmed its contribution to classification.

3.2.2 Model Performance on Prospective Internal and External Validation Sets

The performance of the model trained on Narrow-Band Imaging (NBI) images was first evaluated on a prospective internal validation set collected at PUMC Hospital. The results are summarized in the confusion matrix shown in Figure 4. The model achieved an overall accuracy of 73.44%, with precision of 71.97%, recall of 73.08%, specificity of 86.89%, and F1-score of 71.97%. Misclassified cases were

predominantly those with edge-of-frame lesions, poor texture clarity, or inadequate focal alignment.

The same NBI-trained model was further evaluated on the external POLAR dataset, a multi-center public benchmark comprising pathologically confirmed polyp images. As shown in Figure 5, the confusion matrix illustrates the model's generalizability across institutions. On this external test set, the model achieved an accuracy of 69.46%, with precision of 52.18%, recall of 55.39%, specificity of 82.53%, and F1-score of 53.12%.

The WLI-trained model achieved 57.14% accuracy on prospective internal validation (Figure 6; precision 68.97%, recall 52.63%, specificity 91.74%, F1 59.70% .

However, since the POLAR dataset does not include WLI images, external validation of the WLI-trained model was not feasible.

3.2.3 Binary Classification Performance: Neoplastic vs. Non-Neoplastic Polyps

To evaluate the clinical applicability of the model in distinguishing neoplastic from non-neoplastic polyps, we re-grouped the three-class model by combining conventional adenomas and sessile serrated lesions (SSLs) into a single neoplastic category, while hyperplastic polyps were classified as non-neoplastic.

In the internal validation set, the model achieved an overall accuracy of 91.35%, with precision of 90.74%, recall of 91.63%, specificity of 91.63%, and an F1-score of 90.96% for neoplastic polyp identification.

When applied to the prospective internal validation cohort, the model maintained robust performance with an accuracy of 78.13%, with precision of 74.13%, recall of 76.08%, specificity of 76.08%, and an F1-score of 74.88%.

In external validation using the POLAR dataset, the model demonstrated consistent classification ability, yielding an accuracy of 82.49%, with precision of 62.62%, recall of 64.49%, specificity of 64.49%, and an F1-score of 63.42% for binary neoplasia detection.

The WLI-trained model similarly showed binary classification accuracy of 88.26% on internal validation and 81.63% on prospective internal validation (detailed metrics in supplementary data).

3.2.4 Evaluation of the Model's Predictive Success Rate

The model successfully identified the location of all polyps and provided an optical

diagnosis in 100% of the cases. Among these, the proportion of predictions made with high confidence was 92%. Analysis of the images for which the model did not produce high-confidence predictions revealed that contributing factors included poor image quality, motion blur, suboptimal lighting.

4. Discussion

Our multicenter CADx study addresses a critical gap in SSL recognition by training on a balanced, histology-confirmed dataset across all polyp sizes. Context-enriched ROIs (1.5–2× of lesion diameter) consistently improved SSL classification. However, performance degradation on external datasets highlights persistent inter-institutional domain shifts. These findings suggest that domain adaptation and site-specific fine-tuning are essential for deployment, while current performance should support clinical decision-making rather than replace real-time endoscopic assessment.

Most existing AI models for colonoscopic polyp classification are limited by binary or coarse classification frameworks that fail to accurately differentiate SSLs from hyperplastic polyps¹⁹. Even studies incorporating SSLs often grouped all serrated lesions together or suffered from severe SSL underrepresentation, restricting diagnostic granularity and generalizability^{8,15,20,21}. Moreover, prior systems mainly focused on diminutive polyps (≤ 10 mm), reducing their relevance to real-world practice where larger serrated lesions are common and morphologically diverse¹⁹. Our

study was designed to overcome these limitations by training a fully size-inclusive, class-balanced model targeting histologically confirmed SSLs, hyperplastic polyps, and conventional adenomas, thereby enhancing detection capability for clinically significant serrated lesions.

The reduced performance observed during external validation likely reflects domain shift between institutions. Differences in endoscope manufacturers, imaging parameters, and post-processing settings can alter image appearance and reduce model stability when applied across centers^{22,23}. Recent work has demonstrated the value of domain adaptation—such as deep reconstruction-recoding network (DRRN) architectures that enhance cross-site robustness without requiring new annotations²³—and federated learning frameworks that allow decentralized optimization while preserving data privacy^{24,25}. Domain adaptation and institution-specific fine-tuning to maintain consistent clinical performance.

Our study demonstrated that expanding the ROI to include 1.5–2× the lesion area significantly improved model performance. This contextual approach contrasts with traditional studies that used the smallest bounding boxes around lesion boundaries^{26–28}. However, our results suggest that including contextual information from surrounding normal tissue enhances diagnostic accuracy, particularly for challenging lesions such as sessile serrated lesions (SSLs).

The improved performance with expanded ROI likely reflects that surrounding normal mucosa provides contextual reference for relative tissue comparisons and captures vascular and surface patterns beyond the lesion margin that aid histological prediction^{25,29,30}.

Prior approaches using strict pixel-level segmentation within lesion boundaries struggled with subtle SSLs³¹. Our findings suggest that future AI development should consider incorporating contextual information, particularly for systems aimed at detecting and classifying serrated lesions.

This study has limitations that should guide future work. Our model has not yet been evaluated in prospective randomized clinical trials, which are necessary to determine real-world impact on decision-making and outcomes. Although SSL representation was substantially expanded, performance remained >80% internally but only ~70% externally, underscoring inadequate generalizability for broad deployment and the need for stronger domain adaptation and data augmentation strategies²⁰. Ensemble integration with clinical risk factors³² may further enhance diagnostic robustness.

The ongoing prospective cohort (NCT06773832) is designed to power a subsequent multicenter RCT evaluating SERRATE as a real-time decision-support overlay against standard endoscopist optical diagnosis, with SSL miss rate and surveillance

interval accuracy as primary endpoints.

Declarations

Ethics approval and consent to participate

This study was performed in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of Peking Union Medical College Hospital (approval number: K7281). The study comprised two parts with different consent procedures. For the prospective validation cohort, written informed consent was obtained from all enrolled participants. For the retrospective image datasets, which involved analysis of previously collected, de-identified colonoscopy images, the Ethics Committee of Peking Union Medical College Hospital waived the requirement for informed consent owing to the retrospective nature of the study.

Clinical trial number

NCT06773832 (ClinicalTrials.gov). Date of registration: 14 January 2025.

Consent for publication

Not applicable. This manuscript does not contain any individual person's identifiable data.

Availability of data and materials

The colonoscopy image datasets generated and analysed during the current study from Peking Union Medical College Hospital are not publicly available owing to institutional ethics and patient privacy restrictions, but are available from the corresponding author on reasonable request. The POLAR external validation dataset is publicly available, as cited in the Methods. The code and trained model weights for SERRATE are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by funding obtained by D.W., including: the Excellence Clinical Research Program for Research-based Wards (No. BRWEP2024W034010103); the National Key Research and Development Program of China (No. 2024YFA0918504); the Chinese Academy of Medical Sciences (CAMS) Innovation Fund for Medical Sciences (No. 2022-I2M-1-003); the National High Level Hospital Clinical Research Funding (2022-PUMCH-B-024, 2022-PUMCH-A-071); and the Peking Union Medical College Hospital Talent Cultivation Program (Category C) (No. UBJ11356).

Authors' contributions

W.H., E.C., and L.Z. contributed equally to this work. W.H., Y.Y., and D.W. conceived and designed the study. W.H., L.Z., T.H., K.S., Y.W., Lo.Z., P.Y., B.T., M.Z., J.L., S.Z., Q.J., Z.F., N.Y., X.Y., Y.L., M.Y., and Y.Z. collected the colonoscopy images and performed clinical data curation and lesion annotation. Lo.Z. provided pathological confirmation of lesion diagnoses. E.C., Q.C., B.L., and Y.C. developed the deep learning model, implemented the algorithms, and performed the computational analyses. W.H., E.C., and L.Z. analysed and interpreted the results and drafted the manuscript. Y.Y., D.W., and A.Y. critically revised the manuscript for important intellectual content and supervised the study. D.W. obtained funding. All authors read and approved the final manuscript.

References

1. Siegel, R. L., Wagle, N. S., Cercek, A., Smith, R. A. & Jemal, A. Colorectal cancer statistics, 2023. *CA. Cancer J. Clin.* **73**, 233–254 (2023).
2. Morgan, E. *et al.* Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN. <https://doi.org/10.1136/gutjnl-2022-327736> (2023).
3. East, J. E. *et al.* British Society of Gastroenterology position statement on serrated polyps in the colon and rectum. <https://doi.org/10.1136/gutjnl-2017-314005> (2017).
4. Kahi, C. J., Hewett, D. G., Norton, D. L., Eckert, G. J. & Rex, D. K. Prevalence and Variable Detection of Proximal Colon Serrated Polyps During Screening Colonoscopy. *Clin. Gastroenterol. Hepatol.* **9**, 42–46 (2011).
5. Vennelaganti, S. *et al.* Interobserver Agreement Among Pathologists in the Differentiation of Sessile Serrated From Hyperplastic Polyps. *Gastroenterology* **160**, 452–454.e1 (2021).
6. Mori, Y. *et al.* Real-Time Use of Artificial Intelligence in Identification of Diminutive Polyps During Colonoscopy. *Ann. Intern. Med.* **169**, 357–366 (2018).
7. Horiuchi, H. *et al.* Real-time computer-aided diagnosis of diminutive rectosigmoid polyps using an auto-fluorescence imaging system and novel color intensity analysis software. *Scand. J. Gastroenterol.* **54**, 800–805 (2019).
8. Zachariah, R. *et al.* Prediction of Polyp Pathology Using Convolutional Neural Networks Achieves 'Resect and Discard' Thresholds. *Am. J. Gastroenterol.* **115**, 138–

144 (2020).

9. van der Zander, Q. E. W. *et al.* Optical diagnosis of colorectal polyp images using a newly developed computer-aided diagnosis system (CADx) compared with intuitive optical diagnosis. *Endoscopy* **53**, 1219–1226 (2021).
10. Soleymanjahi, S. *et al.* Artificial Intelligence–Assisted Colonoscopy for Polyp Detection. *Ann. Intern. Med.* **177**, 1652–1663 (2024).
11. Makar, J., Abdelmalak, J., Con, D., Hafeez, B. & Garg, M. Use of artificial intelligence improves colonoscopy performance in adenoma detection: a systematic review and meta-analysis. *Gastrointest. Endosc.* **101**, 68–81.e8 (2025).
12. De Lange, G. *et al.* Artificial intelligence for characterization of colorectal polyps: Prospective multicenter study. *Endosc. Int. Open* **12**, E413–E418 (2024).
13. Ultralytics yolov8. 2023.
https://scholar.google.com/citations?view_op=view_citation&hl=en&user=swSrGtsA AAAJ&citation_for_view=swSrGtsAAAAJ:isC4tDSrTZIC.
14. He, K., Gkioxari, G., Dollár, P. & Girshick, R. Mask R-CNN. Preprint at <https://doi.org/10.48550/arXiv.1703.06870> (2018).
15. Houwen, B. B. S. L. *et al.* Computer-aided diagnosis for optical diagnosis of diminutive colorectal polyps including sessile serrated lesions: a real-time comparison with screening endoscopists. *Endoscopy* **55**, 756–765 (2023).
16. Djinbachian, R. *et al.* Autonomous Artificial Intelligence vs Artificial Intelligence–Assisted Human Optical Diagnosis of Colorectal Polyps: A Randomized Controlled Trial. *Gastroenterology* **167**, 392–399.e2 (2024).

17. Corley, D. A. *et al.* Adenoma Detection Rate and Risk of Colorectal Cancer and Death. *N. Engl. J. Med.* **370**, 1298–1306 (2014).
18. Pilonis, N. D. *et al.* Adenoma Detection Rates by Physicians and Subsequent Colorectal Cancer Risk. *JAMA* **333**, 400–407 (2025).
19. Hann, A., Troya, J. & Fitting, D. Current status and limitations of artificial intelligence in colonoscopy. *United Eur. Gastroenterol. J.* **9**, 527–533 (2021).
20. Yoon, D. *et al.* Colonoscopic image synthesis with generative adversarial network for enhanced detection of sessile serrated lesions using convolutional neural network. *Sci. Rep.* **12**, 261 (2022).
21. Schöler, J. *et al.* Impact of AI-aided colonoscopy in clinical practice: a prospective randomised controlled trial. *BMJ Open Gastroenterol.* **11**, (2024).
22. Kim, M. J. *et al.* The Advent of Domain Adaptation into Artificial Intelligence for Gastrointestinal Endoscopy and Medical Imaging. *Diagnostics* **13**, 3023 (2023).
23. Xu, J. *et al.* Deep reconstruction-recoding network for unsupervised domain adaptation and multi-center generalization in colonoscopy polyp detection. *Comput. Methods Programs Biomed.* **214**, 106576 (2022).
24. Rahman, A. *et al.* Federated learning-based AI approaches in smart healthcare: concepts, taxonomies, challenges and open issues. *Clust. Comput.* 1–41 (2022) doi:10.1007/s10586-022-03658-4.
25. Biffi, C. *et al.* REAL-Colon: A dataset for developing real-world AI applications in colonoscopy. *Sci. Data* **11**, 539 (2024).
26. Poon, C. C. Y. *et al.* AI-doscopist: a real-time deep-learning-based algorithm for

localising polyps in colonoscopy videos with edge computing devices. *Npj Digit. Med.* **3**, 73 (2020).

27. Yamada, M. *et al.* Development of a real-time endoscopic image diagnosis support system using deep learning technology in colonoscopy. *Sci. Rep.* **9**, 14465 (2019).
28. Albuquerque, C., Henriques, R. & Castelli, M. A stacking-based artificial intelligence framework for an effective detection and localization of colon polyps. *Sci. Rep.* **12**, 17678 (2022).
29. Kim, E. S. & Lee, K.-S. Artificial intelligence in colonoscopy: from detection to diagnosis. *Korean J. Intern. Med.* **39**, 555–562 (2024).
30. ELKarazle, K., Raman, V., Then, P. & Chua, C. Detection of Colorectal Polyps from Colonoscopy Using Machine Learning: A Survey on Modern Techniques. *Sensors* **23**, 1225 (2023).
31. Kaminski, M. F. *et al.* Performance measures for lower gastrointestinal endoscopy: a European Society of Gastrointestinal Endoscopy (ESGE) quality improvement initiative. *United Eur. Gastroenterol. J.* **5**, 309–334 (2017).
32. Jiang, H. *et al.* Cost-effectiveness analysis of a community-based colorectal cancer screening program in Shanghai, China. *Front. Public Health* **10**, 986728 (2022).

Tables

Table 1. The distribution of the training and validation dataset for CADx model development

Image type	Pathology	Training set	Validation set	Total
White light	Hyperplastic polyps	477	187	664
	Conventional adenomas	1042	367	1409
	Sessile serrated lesions	524	182	706
NBI	Hyperplastic polyps	396	147	543
	Conventional adenomas	205	83	288
	Sessile serrated lesions	294	120	414

Table 2. The distribution of the prospective internal validation dataset

Pathology	White light	NBI	Total
Hyperplastic polyps	38	38	76
Conventional adenomas	54	68	122
Sessile serrated lesions	55	40	95

Table 3. The performance of different AI models on internal dataset

Models	Accuracy%	Precision%	Recall%	Specificity%	F1-Score%
ResNext50	55.48	54.41	54.29	77.14	53.46
EfficientNetb2	66.49	64.57	64.48	82.24	64.34

ResNet50	62.47	59.87	69.92	79.96	58.98
MobileNetv3	54.02	52.07	52.14	76.07	52.06
EfficientNetb3	68.77	67.92	66.49	83.24	66.06
ConvNext	55.09	56.21	54.96	77.48	52.77
Swin Transformer	73.19	71.22	70.51	85.25	70.08

Table 4. The performance of Swin CADx model with different image type and magnification of surrounding mucosa

Image type	Magnification	Accuracy%	Precision%	Recall%	Specificity%	F1-Score%
WLI	1×	67.04	67.73		79.57	68.35
	1.5×	72.71	72.05		86.36	71.96
	2×	71.62	70.71		85.81	70.70
NBI	1×	73.20	75.81	72.60	86.60	73.75
	1.5×	78.96	78.86	86.99	89.48	78.87
	2×	82.42	82.47	95.21	91.21	82.17

Figures

Figure 1. Overall workflow of the SERRATE CADx study pipeline.

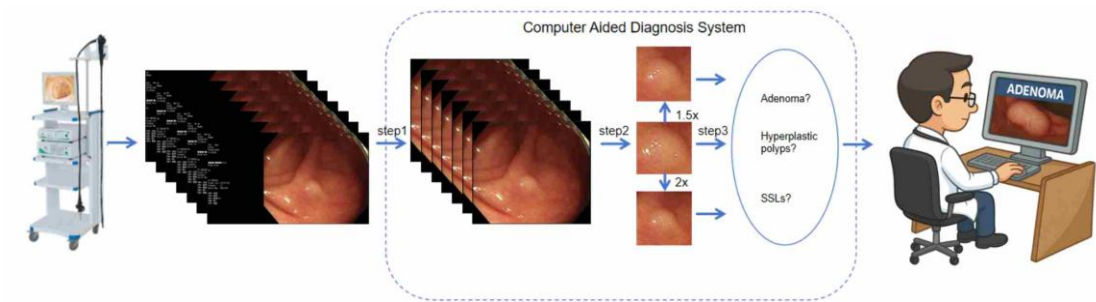
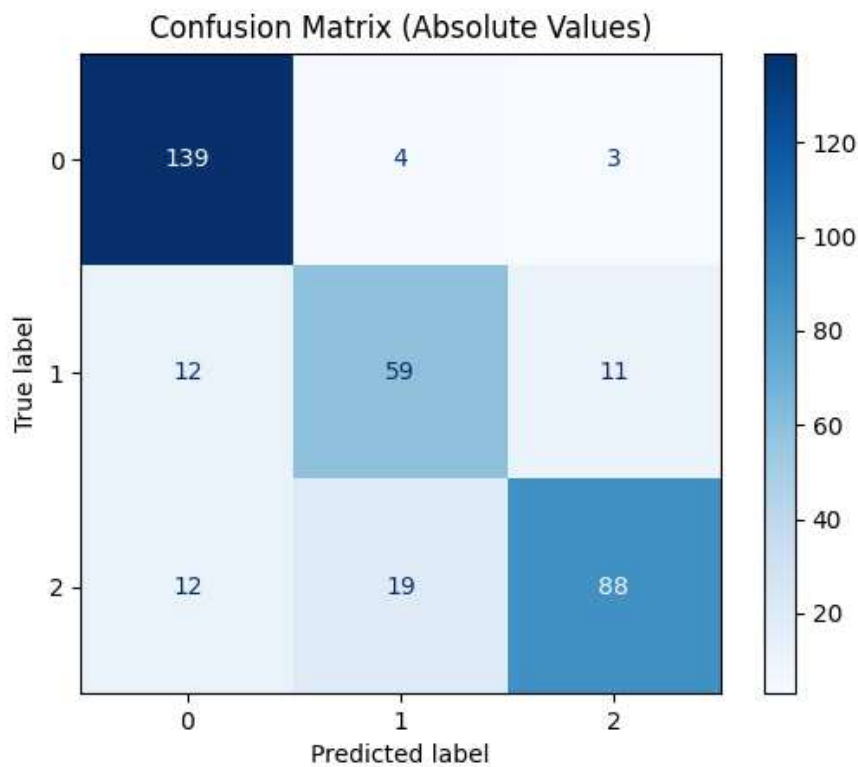


Figure 2. Confusion matrix of the SERRATE model using NBI and 2× context-enriched ROIs on the internal validation set.



Class labels: 0 = hyperplastic polyps, 1 = conventional adenomas, 2 = sessile serrated lesions.

Figure 3. Visualization of SERRATE model predictions across different pathological subtypes: (A) hyperplastic polyps, (B) conventional adenomas, (C) sessile serrated lesions, (D) traditional serrated adenomas.

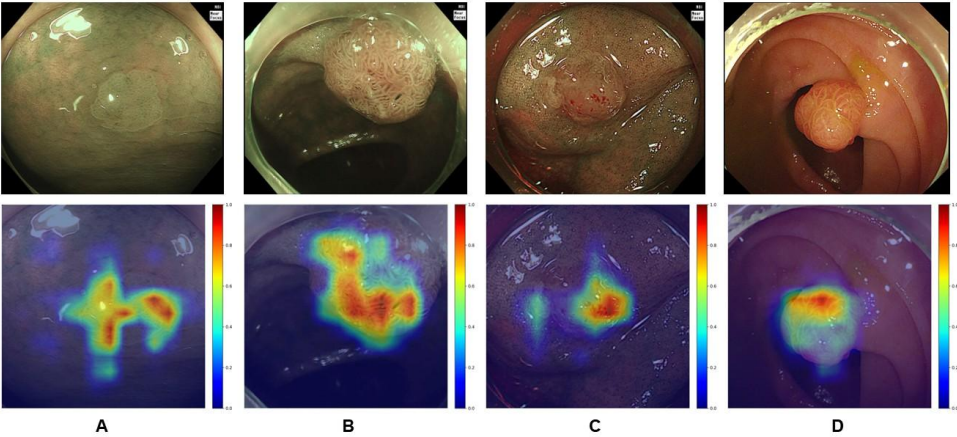


Figure 4. Confusion matrix of the SERRATE model using NBI and 2× context-enriched ROIs on the prospective internal validation set. Class labels as in Figure 2.

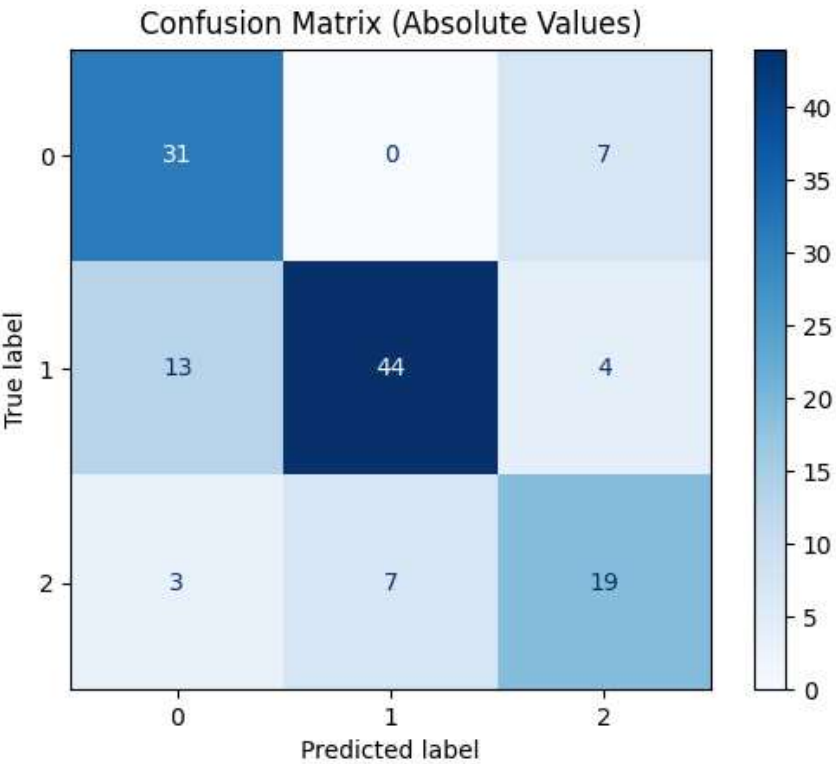


Figure 5. Confusion matrix of the SERRATE model using NBI and 2× context-enriched ROIs on the external multicenter POLAR dataset. Class labels as in Figure 2.

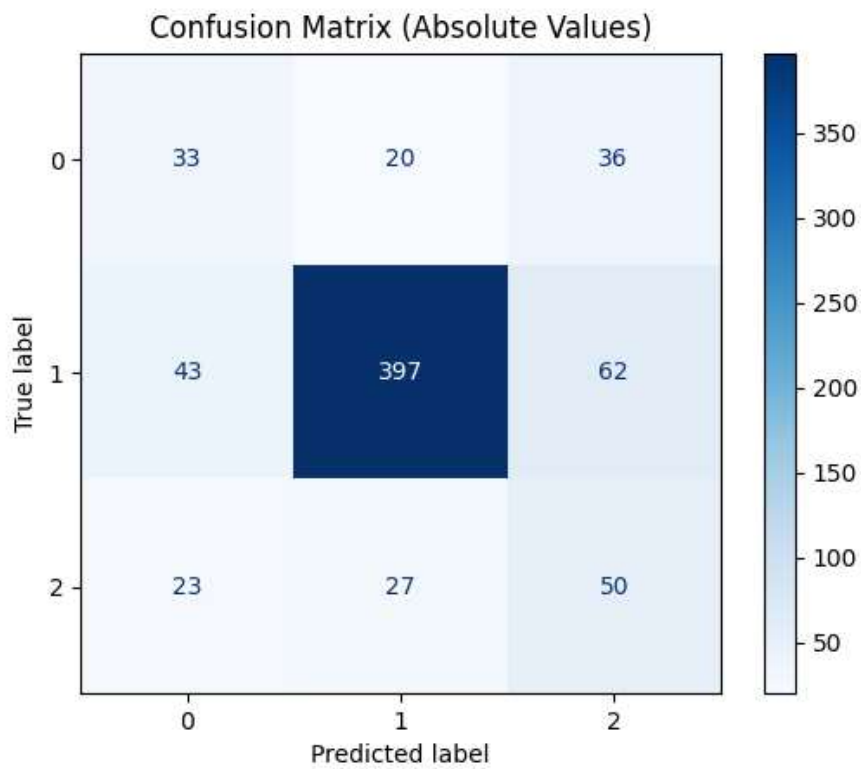


Figure 6. Confusion matrix of the SERRATE model using white-light imaging and 2× context-enriched ROIs on the prospective internal validation set. Class labels as in Figure 2.

