

Computer-Aided Diagnosis of Serrated Colorectal Lesions Using Non-Magnified White Light Endoscopic Images

Daiki Nemoto

Aizu Medical Center Fukushima Medical University

Zhe Guo

The University of Aizu

Boyuan Peng

The University of Aizu

Ruiyao Zhang

The University of Aizu

Yuki Nakajima

Aizu Medical Center Fukushima Medical University

Yoshikazu Hayashi

Jichi Medical University

Takeshi Yamashina

Kansai Medical University Medical Center

Masato Aizawa

Aizu Medical Center Fukushima Medical University

Kenichi Utano

Aizu Medical Center Fukushima Medical University

Alan Kawarai Lefor

Jichi Medical University

Xin Zhu

The University of Aizu

Kazutomo Togashi (✉ togashik@fmu.ac.jp)

Aizu Medical Center Fukushima Medical University

Research Article

Keywords: colonoscopy, colonic polyps, artificial intelligence, serrated lesion

Posted Date: April 14th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1542284/v1>

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

[Read Full License](#)

Abstract

Purpose: Computer-aided diagnosis systems for polyp characterization are commercially available but cannot recognize subtypes of sessile lesions. This study aimed to develop a computer-aided diagnosis system to characterize polyps using non-magnified white-light endoscopic images.

Methods: A total of 2249 non-magnified white-light images from 1030 lesions including 534 tubular adenomas, 225 sessile serrated adenoma/polyps and 271 hyperplastic polyps in the proximal colon were consecutively extracted from an image library and divided into training and testing datasets (4:1), based on the date of colonoscopy. Using ResNet-50 networks, we developed a classifier (1) to differentiate adenomas from serrated lesions, and another classifier (2) to differentiate sessile serrated adenoma/polyps from hyperplastic polyps. Diagnostic performance was assessed using the testing dataset. The computer-aided diagnosis system generated a probability score for each image, and a probability score for each lesion was calculated as the weighted mean with a \log_{10} -transformation. Two experts (E1, E2) read the identical testing dataset with a probability score.

Results: The area under the curve of classifier (1) for adenomas was equivalent to E1 and superior to E2 (classifier 86%, E1 86%, E2 69%; classifier vs. E2, $p < 0.001$). In contrast, the area under the curve of classifier (2) for sessile serrated adenoma/polyps was inferior to both experts (classifier 55%, E1 68%, E2 79%; classifier vs. E2, $p < 0.001$).

Conclusion: The classifier (1) developed using white-light images alone compares favorably with experts in differentiating adenomas from serrated lesions. However, the classifier (2) to identify sessile serrated adenoma/polyps is inferior to experts.

Introduction

Colorectal carcinogenesis through a serrated pathway has been shown, and it is widely accepted that 20–30% of colorectal cancers originated from serrated lesions.[1, 2] According to the World Health Organization (WHO) 2019 classification [3], serrated colorectal lesions are classified into hyperplastic polyps, sessile serrated lesions (SSLs) and traditional serrated adenomas. In the proximal colon, SSLs which were previously referred to as sessile serrated adenomas/polyps (SSAPs) in the WHO 2010 classification [4] play a major role in this serrated pathway. Colonoscopic examination of serrated lesions reveals flat, mucus-covered, pale lesions, with wide variation among colorectal polyps.[5] As for small (≤ 10 mm) polyps, endoscopic differentiation of serrated lesions from conventional adenomas is a challenge for colonoscopists.[6] In previous reports [7–9], image-enhanced endoscopy, e.g., narrow band imaging (NBI) or blue laser imaging facilitated their differentiation, and the diagnostic ability of imaging modalities was excellent to discriminate serrated lesions from conventional adenomas. In contrast, the diagnostic ability using white light images (WLIs) alone was not acceptable, [10, 11] although the overwhelming majority of colonoscopy practice in community hospitals is performed with WLIs and ordinary colonoscopes.

With progress in computer science, artificial intelligence (AI) is being increasingly applied to the interpretation of medical images. In the field of colonoscopy, computer-aided diagnosis (CADx) systems for polyp characterization have been developed and their ability validated. [12–15] To obtain high diagnostic accuracy for real-time prediction of pathology, unfortunately, these CADx systems function only using enhanced images, i.e., NBI [13, 14], or endocytoscopic images [15]. In a recent report, however, the CADx system developed by Zachariah et al showed excellent diagnostic accuracy among polyps, independent of the use of NBI or WLI.[16] This report implies that a CADx for polyp characterization could be developed without magnification and/or image-enhancement. In the present study, we developed CADx systems for polyp characterization, using non-magnified WLI alone. A classifier to differentiate conventional adenomas from serrated lesions was developed initially. In the next step, a classifier to differentiate SSAPs from hyperplastic polyps was developed, which is a difficult problem to solve because even expert endoscopists cannot accurately predict the histology of serrated lesions when using magnified NBI. In the final step, the diagnostic ability of these two classifiers was examined and directly compared with that of expert colonoscopists.

Methods

Study Design

We conducted a single-center, retrospective study. This study was approved by the Institutional Review Board of Fukushima Medical University (registration No. 2020-039). All data were collected by September 2020. The STROBE (Strengthening the reporting of observational studies in epidemiology) guidelines were followed in reporting this study.

Endoscopic Images

To protect patients' privacy, only endoscopic still images were extracted, which were all deidentified and labeled with basic endoscopic information, from existing image libraries in the medical information system of Aizu Medical Center Hospital. Lesions from the images collected were limited to tubular adenomas (TAs) with low-grade dysplasia, SSAPs, and microvesicular type hyperplastic polyps (MVHPs), all located in the proximal colon (cecum, ascending and transverse colon). Images with invasive cancers, villous adenomas, adenomas with high-grade dysplasia, and other miscellaneous lesions were excluded. The images were limited to non-magnified WHIs, and any images enhanced with dye spray, NBI or, blue laser imaging were excluded. Lesions with pedunculated morphology were also excluded. Basic information regarding the lesions included pathological diagnosis (TA, SSAP, MVHP), size and morphology (flat or polypoid). Pathological diagnosis was made by one expert pathologist (Professor Hiroshi Hojo), based on the WHO criteria (2010) [4]. Morphology was classified into polypoid (0-Is), and flat types (0-IIa, 0-IIa + IIc, 0-IIc) based on the Paris classification¹⁷. All endoscopic images were digitized at high resolution (1280×1024), using equipment from two endoscope manufacturers (Fujinon 71% and Olympus 29%).

Allocation into Training and Testing Datasets

A total of 2249 non-magnified WLIs from 1030 lesions were collected including 534 TAs, 225 SSAPs and 271 MVHPs. Characteristics of the lesions are shown in Supplementary Table 1. These images were allocated into a training dataset (with TA 447, SSAP 184, MVHP 184) and a testing dataset (TA 87, SSAP 41, MVHP 87), based on the date of colonoscopy, as illustrated in Fig. 1. In this allocation, lesions detected earlier were used in the training dataset while more recent ones were used in the testing dataset. The ratio of images in the training to testing datasets was set to approach 4:1. Lesion size, anatomical location and endoscope manufacturer were almost evenly distributed in both datasets, but the median number of images per lesion, lesion histology, and morphology showed significant differences between the two datasets (Table 1).

Training Methods

We used a Resnet-50 convolutional neural network as the backbone for extracting features from images for classification [18]. ResNet-50 was pretrained using the ImageNet Large Scale Visual Recognition Challenge 2012 classification dataset, consisting of 1.2 million training images, with 1,000 classes of objects. [19] The CADx systems were proposed and trained using the training dataset. As a first step, we developed a binary classifier (CADx1) to differentiate TAs from serrated lesions. In the next step, we developed another binary classifier (CADx2) to differentiate SSAPs from MVHPs. Before training, lesions were annotated and cropped from endoscopic images for training. Data augmentation including rotation, saturation adjustment, resizing, and exposure adjustment were performed to increase the number of training images. The features of images with lesions were evaluated using ResNet-50, and then sent to a full connection and Focal loss layer for classification.[20] The CADx system was trained using stochastic gradient descent with momentum. In training, the initial learning rate was set to 0.85, a learning rate drop factor 0.88, a minibatch size 30, and a L2 regularization coefficient 0.0005. A focal loss [21] was employed to overcome the influence caused by the imbalance of images' numbers in different classes, where alpha was set to 0.1, gamma was set to 2. The system is developed using MATLAB 2020a (MathWorks Inc., US) and the Deep Learning Toolbox™ and Parallel Computing Toolbox™, and a workstation with two Graphics Processing Units with NVIDIA GeForce GTX TITAN X.

Testing Methods

After the training process, diagnostic performance was assessed using the testing dataset. The CADx1 system generated a probability score of each image for TA, and the probability score for each lesion was calculated as the weighted mean of the \log_{10} -transformed probability scores of images. [22] A probability score > 0.5 was defined as TA. Similarly, the CADx2 system to differentiate SSAP from MVHP was tested.

Readings by Endoscopists

Two expert endoscopists (YH, TY) from other institutions were asked to compare their diagnostic performance with the CADx system. Both experts have performed over 10,000 colonoscopic examinations each and are certified by the Japan Gastroenterological Endoscopy Society. Identical endoscopic images were presented in a random order, and the experts rated the histology with a

confidence level (high, medium, low), blinded to the proportion of each histology. First, experts rated whether the polyp histology is TA or a serrated lesion. Two months later, experts rated whether the histology of serrated lesions is SSL or MVHP. To discriminate TAs from serrated lesions, for instance, an expert diagnosis of TA with low confidence level is interpreted as a “0.6” probability score, and an expert diagnosis of serrated lesions with high confidence is interpreted as a “0” probability score (Supplementary Table 2). After completing the reading test, a probability score for each lesion was calculated as the weighted mean of \log_{10} -transformed probability score of images, the same as CADx.

Outcome Measurements and Statistical Methods

Area under the curve (AUC) analysis from the receiver operating characteristics curve was performed based on the weighted mean probability score of images for each lesion. The primary outcome is AUC for CADx1 to differentiate TAs from serrated lesions. Secondary outcomes are other diagnostic values (sensitivity, specificity and predictive value for CADx1, AUC of CADx2 for SSAP, and diagnostic values for the histology of each lesion. Each value is expressed as a percentage [95% confidence intervals (CI)]. The Clopper-Pearson method was used for assessing the 95% CI. STARD (Standards for Reporting of Diagnostic Accuracy Studies) recommendations were followed in reporting this study. The chi-square test was used to compare AUC and diagnostic values between the CAD system and experts. Subgroup analyses by lesion size, morphology, and endoscope manufacturer were also performed. All p-values are 2-tailed, and values < 0.05 indicate statistical significance. All statistical analyses were performed with Stata 16 (Stata Corp., College Station, TX, USA).

Results

Diagnostic Performance to Differentiate TA from Serrated Lesions (Table 2)

The AUC was 86% [95% CI: 80 - 91] in CADx1, 86% [80 - 91] for expert 1, and 69% [62 - 76] for expert 2 (Figure 2A). A significant difference was observed between CADx1 and expert 2 ($p=0.0003$). The accuracy showed a tendency similar to the AUC. Diagnostic performance of CADx1 by probability score is shown in Supplementary Table 3A. The sensitivity of CADx1 was significantly higher than that for both experts while the specificity of CADx1 was significantly lower than that of expert 1. The positive predictive value showed a higher trend for expert 1, and the negative predictive value was significantly lower for expert 2. Examples are shown in Figure 3.

Diagnostic Performance to Differentiate SSAP from MVHP (Table 3)

The AUC was 55% [44 - 66] for CADx2, 68% [58 - 77] for expert 1 and 79% [70 - 87] for expert 2 (Figure 2B). Significant differences were observed between CADx2 and expert 2 ($p=0.0002$). Diagnostic performance of CADx2 by probability score is shown in Supplementary Table 3B. The accuracy showed similar values for all readers. The sensitivity of CADx2 was significantly lower than that of both experts while the specificity of CADx2 was significantly higher than that of both experts. The positive predictive value

showed no apparent trend. The negative predictive value of CADx2 showed a lower trend, and a significant difference was observed between CADx2 and expert 2. Examples are shown in Figure 4.

Subgroup Analyses (Table 4)

In differentiating TA from serrated lesions, there is a trend that the diagnostic performance for larger (≥ 2 cm) lesions was superior to that for smaller lesions for all readers, and this trend is apparent in the AUC of CADx1. In subgroup analyses by morphology, the AUC of CADx1 shows a higher trend for polypoid lesions, but this trend was not observed in the accuracy. In differentiating SSAP from MVHP, the diagnostic performance for larger (≥ 2 cm) lesions was inferior for CADx2, and this trend was statistically significant for accuracy ($p= 0.008$). In contrast, these trends were not observed for both experts. A statistical comparison of AUC was not possible based on morphology because all 8 polypoid lesions were MVHP, and comparison of the accuracy did not show a trend. Subgroup analysis by endoscope manufacturer is shown in Supplementary Table 4. The AUC of CADx2 for “Fujifilm” was significantly higher than that for “Olympus” ($p=0.001$).

Diagnostic Performance based on image type

The overall diagnostic performance of CADx1 for TA versus serrated lesions and CADx2 for SSAP versus MVHP are summarized in Supplementary Table 5.

Discussion

This study demonstrates that the diagnostic ability of the CADx1 system to differentiate TA (tubular adenomas with low-grade dysplasia) from serrated lesions (SSAP + MVHP) is superior or equivalent to experts. These excellent results may be attributed to the collection of a relatively large number of endoscopic images of serrated lesions including SSAPs for deep learning. Application of the latest AI technology such as focal loss and over-sampling may overcome imbalances in the data resulting in good diagnostic performance. [20, 21] In subgroup analysis, the AUC of the CADx1 increased with lesion size, reaching as high as 94% [88–100] in the largest (> 10 mm) lesion group. The AUC was also as high as 94% [87–100] for polypoid lesions. The present CADx1 system may be practical for lesion classification in clinical use, particularly for large polypoid lesions.

In contrast, the diagnostic ability of the CADx2 system to differentiate SSAP from MVHP is inferior to both experts, although the diagnostic ability of the experts was not acceptably high. In previous reports, expert endoscopists were unable to make accurate optical diagnoses for the subtype of serrated lesions even using magnified NBI. [23] Since it is generally difficult for CADx to surpass human diagnostic abilities with current technology, these poor results might be within expectations. Development of a CADx system to accurately characterize lesions including the subtype of serrated lesions is a technological challenge, and most previous studies did not include serrated lesions in either training or testing datasets. [11–15] Unstable histological assessment would influence these less than acceptable results. In a recent report, the overall interobserver agreement for serrated lesions was moderate ($\kappa = 0.44$) based on

histological assessment. [24] Recent minor changes in the histological criteria for SSL, where some MVHPs are reclassified into SSLs, would indicate that the SSL is not a distinct disease entity. Recent evidence has established that MVHP is a precursor to SSL, suggesting that both lesions are along the same disease spectrum. In clinical practice, the presence of a lesion > 1cm in the proximal colon is an indication for polypectomy regardless of histology. Except for SSLs, small (< 1cm) serrated lesions are not good candidates for endoscopic resection. Accuracy of the CADx2 in the smallest lesion group revealed the highest (73%) accuracy and approaches the diagnostic ability of experts. Application of the CADx2 system might be limited to small and polypoid lesions.

To evaluate the diagnostic performance of the CADx system, we selected the AUC as the primary outcome because comparisons are easily made regardless of discrimination thresholds. For each image, the CADx system generates a probability score which expresses its prediction of lesion histology, and we set 0.5 as the discrimination threshold in this study. However, this discrimination threshold can be arbitrarily set to obtain the highest diagnostic value. At a discrimination threshold of 0.5, the sensitivity of the CADx2 system was extremely low (17%) and the specificity was high (85%), implying that probability score threshold was not optimal. To classify serrated lesions more accurately, lower thresholds may be optimal.

This study has several acknowledged limitations. First, this is a retrospective study, and a prospective clinical trial should be performed to confirm the performance and reliability of the CADx systems. Second, the datasets were extracted based on predefined inclusion/exclusion criteria, but mainly using subjective assessment of image quality, which may lead to selection bias. Third, the training data were still images selected by endoscopists. This limited the full use of deep neural networks to extract features for classification in the training phase. A future study should use digital videos incorporating temporal images with sequence recognition. Fourth, all data are from a single medical center. The proposed method should be trained and tested using data from multiple institutions to improve performance and reliability.

In conclusion, we have developed two CADx systems trained with non-magnified WLIs. The CADx1 was comparable to experts in differentiating TAs from serrated lesions in the proximal colon. The CADx2 system failed to demonstrate sufficient diagnostic ability in discriminating SSAPs from MVPs. This is the next issue to be addressed. A future study may use digital video images incorporating temporal images with sequence recognition.

Declarations

Acknowledgement

The Statistical Consulting Service at Health Data Science Research Section, Tokyo Metropolitan Institute of Gerontology, and Clinical Research Center at Fukushima Medical University, in particular Dr. Noriko Tanaka, helped with the interpretation of the results from statistical analysis in this study. Dr. Hiroshi

Hojo, former professor at Aizu Medical Center Fukushima Medical University helped with the interpretation of the histological assessment.

Conflict of Interests

Authors declare no conflict of interests for this article.

Authors' contributions

Study conception and design: Nemoto, Guo, Hayashi, Utano, Aizawa, Zhu, Togashi

Acquisition of data: Nemoto, Aizawa, Utano, Nakajima, Hayashi, Yamashina, Togashi

Interpretation of data: Guo, Peng, Zhang, Zhu, Nemoto, Togashi

Drafting of the manuscript: Nemoto, Guo, Zhu, Togashi

Critical revision of the manuscript for important intellectual content: Nemoto, Guo, Peng, Zhang, Zhu, Aizawa, Nakajima, Utano, Hayashi, Yamashina, Hewett, Togashi

Statistical analysis: Guo, Zhu, Nemoto, Togashi

Obtained funding: Zhu, Togashi

Administrative, technical, or material support: Zhu, Togashi

Study supervision: Zhu, Togashi

Final manuscript approval: Nemoto, Guo, Peng, Zhang, Aizawa, Nakajima, Utano, Hayashi, Yamashina, Lefor, Zhu, Togashi

Funding Declaration

This work was partially funded by MEXT/JSPS KAKENHI Grant Number JP 18K08010 and the Competitive Research Fund of the University of Aizu Grant Number 2019-P-2.

References

1. Bettington M, Walker N, Clouston A, et al (2013) The serrated pathway to colorectal carcinoma: current concepts and challenges. *Histopathology* 62:367–86. doi: 10.1111/his.12055
2. Rosty C, Hewett DG, Brown IS, et al (2013) Serrated polyps of the large intestine: current understanding of diagnosis, pathogenesis, and clinical management. *J Gastroenterol* 48:287–302. doi: 10.1007/s00535-012-0720-y
3. Pai RK, Mäkinen MJ, Rosty C (2019) Colorectal serrated lesions and polyps. In: WHO Classification of Tumours Editorial Board (ed). *Digestive system tumours*. IARC: Lyon, France, pp 163–70.

4. Snover DC, Ahnen DJ, Burt RW, et al (2010) Serrated polyps of the colon and rectum and serrated polyposis. In: Bosman FT, Carneiro F, Hruban RH, Theise ND, editors. WHO classification of tumours of the digestive system. IARC: Lyon, France, pp 160–165
5. Kashida H. Endoscopic diagnosis of sessile serrated polyp: A systematic review (2019) *Dig Endosc* 31:16–23. doi: 10.1111/den.13263
6. I. Jespeert JE, Bastiaansen BA, van Leerdam ME, et al (2016) Development and validation of the WASP classification system for optical diagnosis of adenomas, hyperplastic polyps and sessile serrated adenomas/polyps. *Gut* 65:963–70. doi: 10.1136/gutjnl-2014-308411
7. Kobayashi S, Yamada M, Takamaru H, et al (2019) Diagnostic yield of the Japan NBI Expert Team (JNET) classification for endoscopic diagnosis of superficial colorectal neoplasms in a large-scale clinical practice database. *United European Gastroenterol J* 7:914–923. doi: 10.1177/2050640619845987
8. Glover B, Teare J, Patel N (2019) A Review of new and emerging techniques for optical diagnosis of colonic polyps. *J Clin Gastroenterol* 53:495–506. doi: 10.1097/MCG.0000000000001222.
9. Togashi K, Nemoto D, Utano K, et al (2016) Blue laser imaging endoscopy system for the early detection and characterization of colorectal lesions: a guide for the endoscopist. *Therap Adv Gastroenterol* 9:50–6. doi: 10.1177/1756283X15603614
10. Rastogi A, Keighley J, Singh V, et al (2009) High accuracy of narrow band imaging without magnification for the real-time characterization of polyp histology and its comparison with high-definition white light colonoscopy: a prospective study. *Am J Gastroenterol* 104:2422–30. doi: 10.1038/ajg.2009.403
11. Ignjatovic A, East JE, Guenther T, et al (2011) What is the most reliable imaging modality for small colonic polyp characterization? Study of white-light, autofluorescence, and narrow-band imaging. *Endoscopy* 43:94–9. doi: 10.1055/s-0030-1256074
12. Komeda Y, Handa H, Watanabe T, et al (2017) Computer-Aided Diagnosis Based on Convolutional Neural Network System for Colorectal Polyp Classification: Preliminary Experience. *Oncology* 93 Suppl 1:30–34. doi: 10.1159/000481227
13. Chen PJ, Lin MC, Lai MJ, et al (2018) Accurate Classification of Diminutive Colorectal Polyps Using Computer-Aided Analysis. *Gastroenterology* 154:568–575. doi: 10.1053/j.gastro.2017.10.010
14. Byrne MF, Chapados N, Soudan F, et al (2019) Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. *Gut* 68:94–100. doi: 10.1136/gutjnl-2017-314547
15. Kudo SE, Misawa M, Mori Y, et al (2020) Artificial intelligence-assisted system improves endoscopic identification of colorectal neoplasms. *Clin Gastroenterol Hepatol* 18:1874–1881.e2. doi: 10.1016/j.cgh.2019.09.009.
16. Zachariah R, Samarasena J, Luba D, et al (2020) Prediction of polyp pathology using convolutional neural networks achieves "resect and discard" thresholds. *Am J Gastroenterol* 115:138–144. doi: 10.14309/ajg.0000000000000429

17. No author (2003) The Paris endoscopic classification of superficial neoplastic lesions: esophagus, stomach, and colon: November 30 to December 1, 2002. *Gastrointest Endosc* 2003;58:S3-43. doi: 10.1016/s0016-5107(03)02159-x
18. Kaiming H, Zhang X, Ren S et al (2016) "Deep residual learning for image recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778.
19. Deng J, Socher R, Li L et al (2009) ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*.
20. Guo Z, Zhang R, Li Q et al (2020) Reduce False-Positive Rate by Active Learning for Automatic Polyp Detection in Colonoscopy Videos. *Proc. IEEE International Symposium on Biomedical Imaging (ISBI'20)*.
21. Lin TY, Goyal P, Girshick R et al (2020) Focal Loss for Dense Object Detection. *IEEE Trans Pattern Anal Mach Intell* 42:318–27. doi: 10.1109/TPAMI.2018.2858826
22. Zhou D, Tian F, Tian X et al (2020) Diagnostic evaluation of a deep learning model for optical diagnosis of colorectal cancer. *Nat Commun* 11:2961. doi: 10.1038/s41467-020-16777-6
23. Parikh ND, Chaptini L, Njei B et al (2016) Diagnosis of sessile serrated adenomas/polyps with image-enhanced endoscopy: a systematic review and meta-analysis. *Endoscopy* 48:731–9. doi: 10.1055/s-0042-107592
24. Vennelaganti S, Cuatrecasas M, Vennalaganti P et al (2021) Interobserver agreement among pathologists in the differentiation of sessile serrated from hyperplastic polyps. *Gastroenterology* 160:452-454e1. doi: 10.1053/j.gastro.2020.09.015.

Tables

Table 1. Characteristics of lesions in the training and testing datasets

Dataset		All	Training	Testing	p-value*
Total number of lesions		1030	815	215	
Total number of images		2249	1849	400	
Number of images per lesion	Median [IQR]	2 [1 - 3]	2 [1 - 3]	1 [1 - 2]	0.0034
Histology	TA, low-grade, n (%)	534 (52%)	447 (55%)	87 (40%)	<0.001
	Serrated lesions**, n (%)	496 (48%)	368 (45%)	129 (60%)	
	SSAP, n (%)	225 (22%)	184 (23%)	41 (19%)	
	MVHP, n (%)	271 (26%)	184 (23%)	87 (40%)	
Size, mm	Median [IQR]	7 [5-10]	7 [5-10]	7 [5-9]	0.20
	<6mm, n (%)	312 (30%)	237 (29%)	75 (35%)	0.18
	6-10mm, n (%)	549 (53%)	438 (54%)	111 (52%)	
	>10mm, n (%)	169 (16%)	140 (17%)	29 (13%)	
Morphology***	Flat, n (%)	727 (71%)	557 (68%)	170 (79%)	0.002
	Polypoid, n (%)	303 (29%)	258 (32%)	45 (21%)	
Anatomical location	Cecum, n (%)	194 (19%)	145 (18%)	49 (23%)	0.16
	Ascending colon, n (%)	395 (38%)	311 (38%)	84 (39%)	
	Transverse colon, n (%)	441 (43%)	359 (44%)	82 (38%)	
Endoscope manufacturer	Fujifilm, n (%)	732 (71%)	585 (72%)	147 (68%)	0.33
	Olympus, n (%)	298 (29%)	230 (28%)	68 (32%)	

IQR: interquartile range; TA: tubular adenoma; SSAP: sessile serrated adenoma/polyp; MVHP: microvesicular type hyperplastic polyp

*p-values for differences between the training and testing datasets were calculated using the Mann-Whitney U test or chi-square test as applicable.

**serrated lesions include SSAP and MVHP

***Morphology is defined according to Paris' classification. Polypoid includes 0-Is whereas flat includes 0-IIa, 0-IIa+IIc and 0-IIc

Table 2. Diagnostic performance differentiating tubular adenomas from serrated lesions

Reader	CADx1	Expert 1	Expert 2	p-value*	
				CADx1 versus Expert 1	CADx1 versus Expert 2
AUC, % [95%CI]	86 [80-91]	86[80-91]	69 [62-76]	0.97	0.0003
Accuracy, % [95%CI]	82 [77-87]	80 [74-85]	65 [58-71]	0.54	<0.001
Sensitivity, % [95%CI]	72 [62-81]	56 [45-67]	23 [15-33]	0.027	<0.001
Specificity, % [95%CI]	89 [82-94]	96 [91-99]	94 [88-97]	0.032	0.18
PPV, % [95%CI]	82 [71-90]	91 [80-97]	71 [51-87]	0.15	0.25
NPV, % [95%CI]	83 [75-89]	76 [69-83]	64 [57-71]	0.19	<0.001

CADx1: computer-aided diagnosis; AUC: area under curve; CI: confidence interval

PPV: positive predictive value; NPV: negative predictive value

*p-values are calculated using chi-square test.

Table 3. Diagnostic performance differentiating SSAP from MVHP

Reader	CADx2	Expert 1	Expert 2	p-value*	
				CADx2 versus Expert 1	CADx2 versus Expert 2
AUC, % [95%CI]	55 [44-66]	68 [58-77]	79 [70-87]	0.077	0.0002
Accuracy, % [95%CI]	63 [54-72]	68 [59-76]	69 [60-77]	0.43	0.36
Sensitivity, % [95%CI]	17 [7-32]	66 [49-80]	88 [74-96]	<0.001	<0.001
Specificity, % [95%CI]	85 [76-92]	69 [58-78]	60 [49-70]	<0.001	0.012
PPV, % [95%CI]	35 [15-59]	50 [36-64]	51 [39-63]	0.25	0.21
NPV, % [95%CI]	69 [59-77]	81 [70-89]	91 [81-97]	0.059	0.001

SSAP: sessile serrated adenoma/polyp; MVHP: microvesicular hyperplastic polyp

CADx2: computer-aided diagnosis; CI: confidence interval

PPV: positive predictive value; NPV: negative predictive value; AUC: area under curve

*p-values are calculated using chi-square test.

Table 4. Diagnostic performance by size and morphology

	Reader	Size			p-value	Morphology*		
		<6mm	6-10mm	>10mm		Polypoid	Flat	p-value
TA vs. Serrated lesions**								
AUC, % [95%CI]	CADx1	79 [67-91]	86 [78-94]	94 [88-100]	0.049	94 [87-100]	84 [77-91]	0.052
	Expert 1	84 [74-93]	85 [77-93]	90 [79-100]	0.67	84 [73-96]	84 [77-91]	0.91
	Expert 2	67 [55-80]	64 [53-76]	84 [71-96]	0.067	58 [41-75]	65 [56-75]	0.48
Accuracy, % [95%CI]	CADx1	80 [69-88]	81 [72-89]	88 [75-95]	0.51	80 [65-90]	83 [76-88]	0.65
	Expert 1	77 [68-87]	78 [70-87]	88 [79-97]	0.30	73 [58-85]	82 [76-88]	0.21
	Expert 2	60 [49-71]	64 [54-74]	76 [63-88]	0.20	47 [32-61]	70 [63-77]	0.003
SSAP vs. MVHP								
AUC, % [95%CI]	CADx2	60 [39-81]	52 [31-72]	47 [26-68]	0.67	n/a***	57 [46-68]	n/a***
	Expert 1	66 [47-85]	63 [45-82]	65 [47-83]	0.99	n/a***	68 [58-78]	n/a***
	Expert 2	75 [54-95]	86 [76-95]	70 [53-87]	0.23	n/a***	78 [69-87]	n/a***
Accuracy, % [95%CI]	CADx2	73 [57-85]	69 [55-81]	41 [24-59]	0.008	63 [24-91]	63 [54-72]	0.10
	Expert 1	77 [65-90]	62 [48-75]	63 [46-79]	0.21	63 [29-96]	68 [59-76]	0.77
	Expert 2	80 [68-91]	65 [52-78]	59 [42-76]	0.14	75 [45-105]	68 [60-77]	0.69

CADx: computer-aided diagnosis; TA: tubular adenoma; CI: confidence interval; AUC: area under curve

SSAP: sessile serrated adenoma/polyp; MVHP: microvesicular type hyperplastic polyp; n/a not available

p-values are calculated using the chi-square test

* Morphology is defined by Paris' classification. Polypoid includes 0-Is whereas flat includes 0-IIa, 0-IIa+IIc and 0-IIc

** Serrated lesions include SSL and MVHP

***Unable to calculate AUC because all 8 polypoid lesions were MVHP

Figures

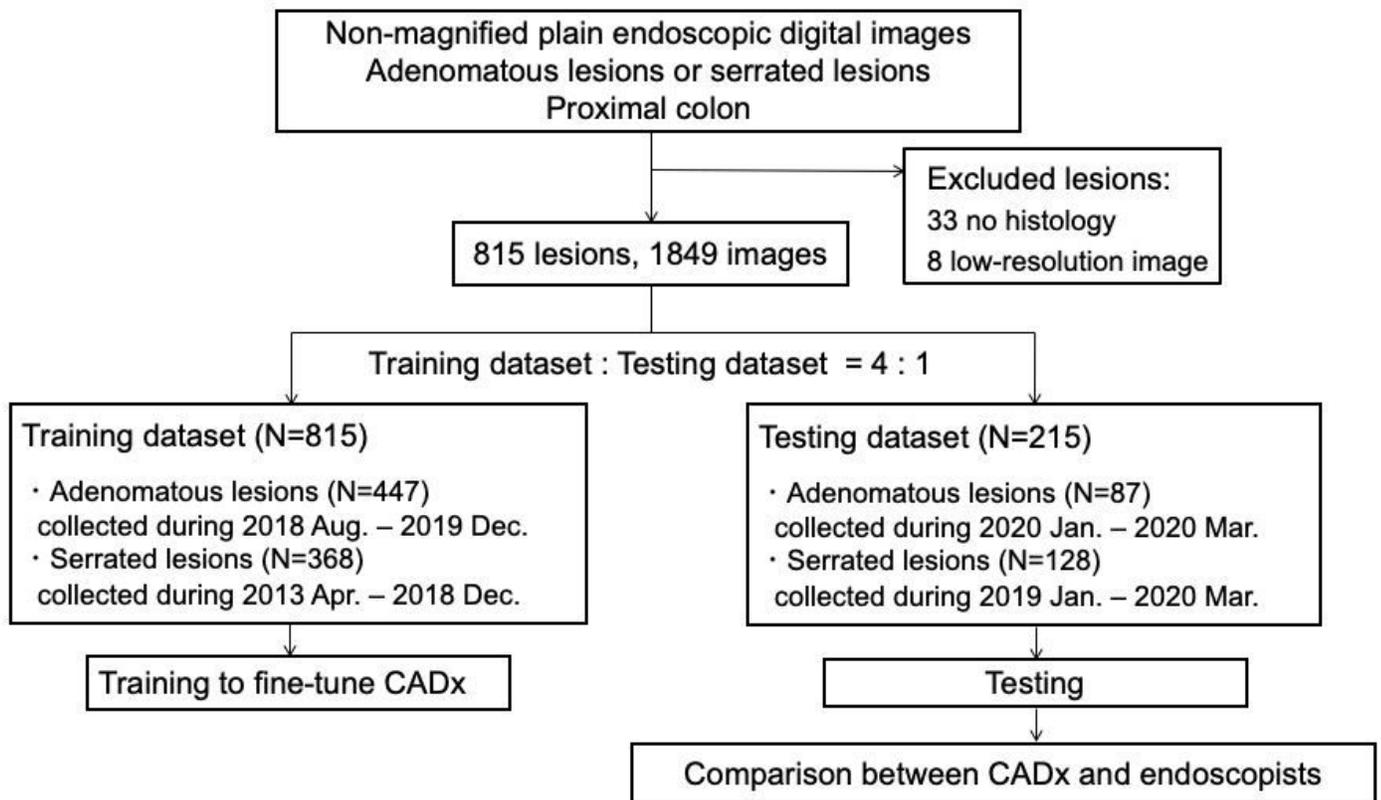


Figure 1

Flowchart of image allocation into training and testing datasets. CADx: computer-aided diagnosis

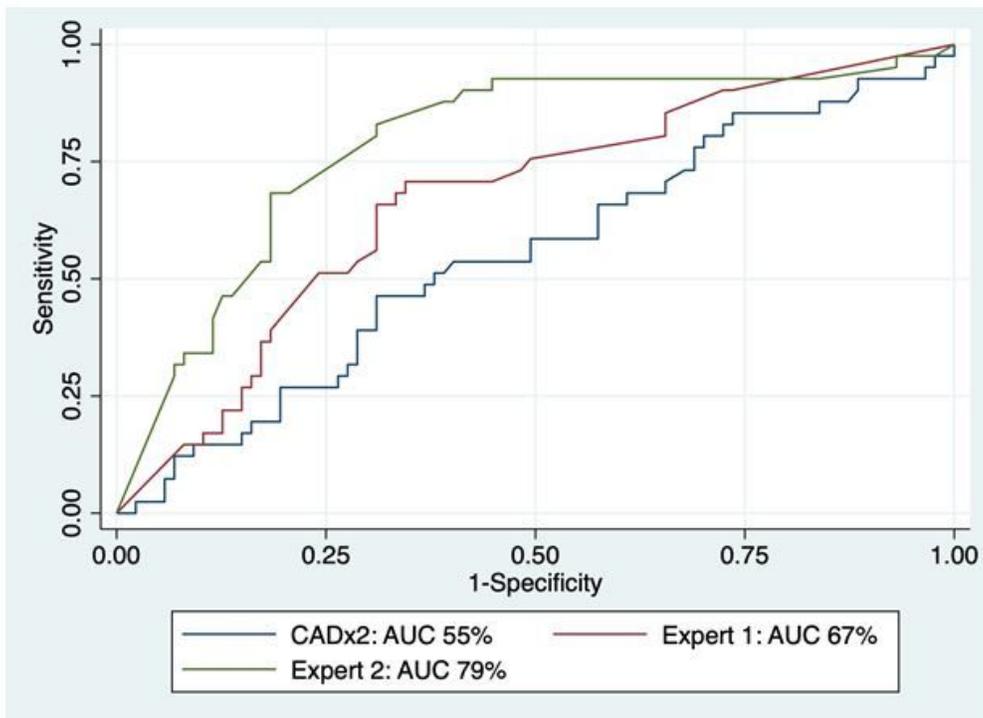
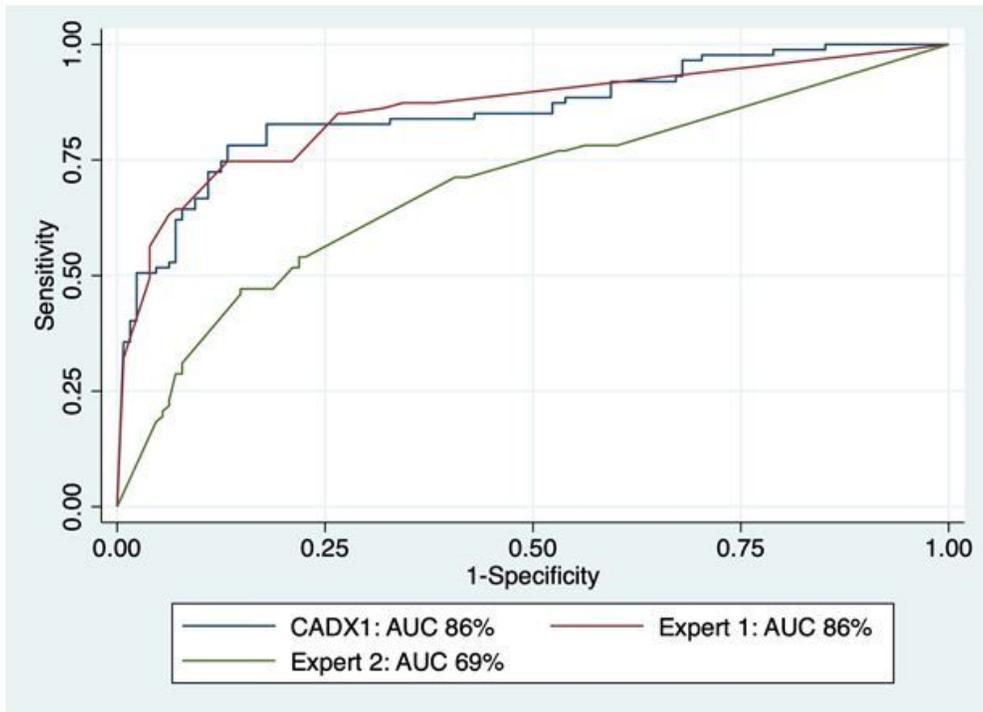


Figure 2

2A. Area under the curve (AUC) in differentiating tubular adenomas from serrated lesions. Receiver operator characteristics curve was calculated based on the weighted mean of \log_{10} -transformed probability scores of images. The AUC was 86% [80-91] for CADx1, 86% [80-91] for expert 1, and 69% [62-76] for expert 2. There is a significant difference between CADx1 and expert 2 ($p=0.0003$).

2B. Area under the curve (AUC) in differentiating sessile serrated adenoma/polyp from microvesicular type hyperplastic polyp. Receiver operator characteristics curve was calculated based on the weighted mean of \log_{10} -transformed probability scores of images. The AUC was 55% [44-66] for CADx2, 68% [58-77] for expert 1, and 79% [70-87] for expert 2. There is a significant difference between CADx2 and expert 2 ($p=0.0002$).

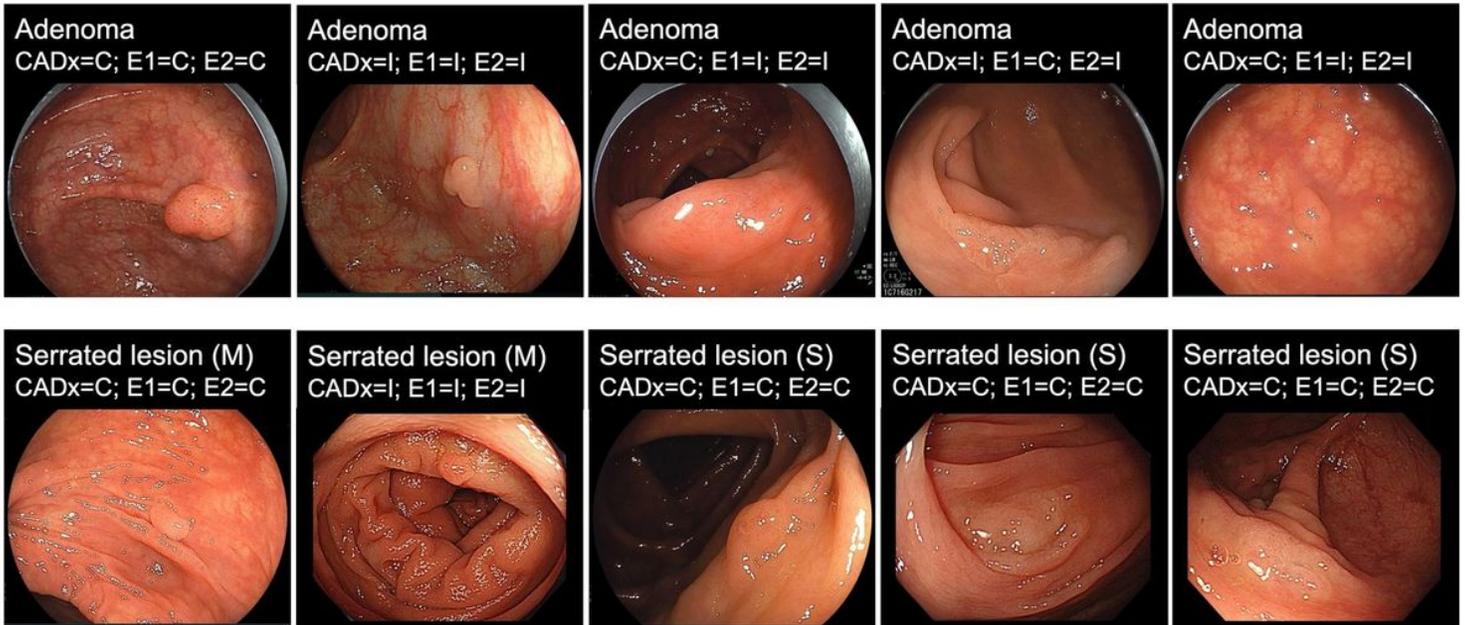


Figure 3

Examples of testing-dataset images for differentiating adenomas from serrated lesions. CADx1: computer-aided diagnosis. E1: expert 1, E2: expert 2, C: correct, I: incorrect, S: sessile serrated lesion, M: microvesicular type hyperplastic polyp

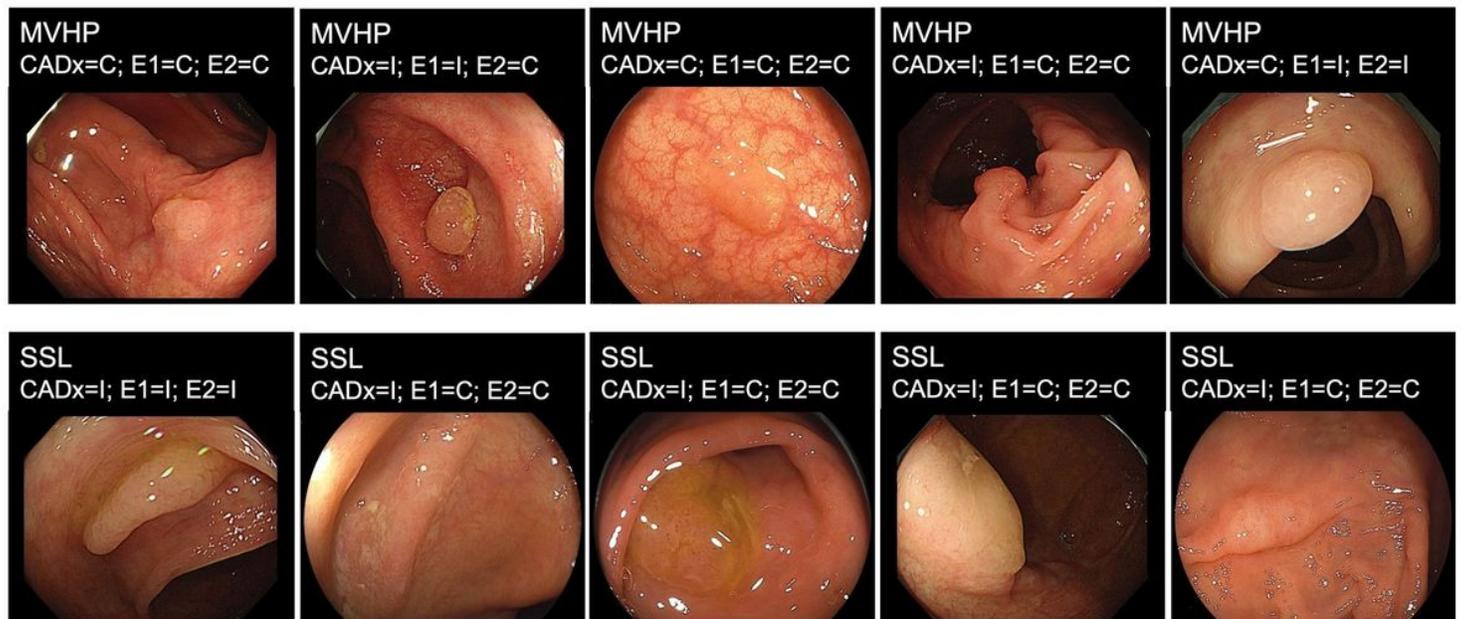


Figure 4

Examples of testing-dataset images for differentiating sessile serrated lesions (SSL) from microvesicular type hyperplastic polyps (MVHP). CADx2: computer-aided diagnosis. E1: expert 1, E2: expert 2, C: correct, I: incorrect.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supple.Table1.STROBE.doc](#)
- [Supple.Table2.docx](#)
- [Supple.Table3.docx](#)
- [Supple.Table4A.docx](#)
- [Supple.Table4B.docx](#)
- [Supple.Table5.docx](#)
- [Supple.Table6.docx](#)