

Automated endoscopic detection and classification of colorectal polyps using convolutional neural networks

Tsuyoshi Ozawa^{ID}, Soichiro Ishihara, Mitsuhiro Fujishiro, Youichi Kumagai, Satoki Shichijo and Tomohiro Tada

Abstract

Background: Recently the American Society for Gastrointestinal Endoscopy addressed the 'resect and discard' strategy, determining that accurate *in vivo* differentiation of colorectal polyps (CP) is necessary. Previous studies have suggested a promising application of artificial intelligence (AI), using deep learning in object recognition. Therefore, we aimed to construct an AI system that can accurately detect and classify CP using stored still images during colonoscopy.

Methods: We used a deep convolutional neural network (CNN) architecture called Single Shot MultiBox Detector. We trained the CNN using 16,418 images from 4752 CPs and 4013 images of normal colorectums, and subsequently validated the performance of the trained CNN in 7077 colonoscopy images, including 1172 CP images from 309 various types of CP. Diagnostic speed and yields for the detection and classification of CP were evaluated as a measure of performance of the trained CNN.

Results: The processing time of the CNN was 20 ms per frame. The trained CNN detected 1246 CP with a sensitivity of 92% and a positive predictive value (PPV) of 86%. The sensitivity and PPV were 90% and 83%, respectively, for the white light images, and 97% and 98% for the narrow band images. Among the correctly detected polyps, 83% of the CP were accurately classified through images. Furthermore, 97% of adenomas were precisely identified under the white light imaging.

Conclusions: Our CNN showed promise in being able to detect and classify CP through endoscopic images, highlighting its high potential for future application as an AI-based CP diagnosis support system for colonoscopy.

Keywords: artificial intelligence, classification, colon, colorectal, convolutional neural network, detection, diagnosis, polyp

Received: 12 October 2019; revised manuscript accepted: 12 February 2020.

Introduction

Colorectal cancer (CRC) is a major public health problem, as the second leading cause of cancer-related death in the United States and the fourth leading cause of cancer-related death worldwide.^{1,2} Approximately 85% of CRC have been suggested to develop from adenomas through genetic and epigenetic changes, and it has been

reported that endoscopic resection of colorectal polyps (CP) reduces the incidence of CRC.^{3,4}

Pathologically, CP are classified into adenoma, hyperplastic polyp, sessile serrated adenoma/polyp (SSAP), and other polyps, such as juvenile and inflammation polyp. The risk of developing CRC is different for each classification.^{3,5} It is

Ther Adv Gastroenterol

2020, Vol. 13: 1–13

DOI: 10.1177/
1756284820910659

© The Author(s), 2020.
Article reuse guidelines:
sagepub.com/journals-
permissions

Correspondence to:

Tsuyoshi Ozawa
Department of Surgery,
Teikyo University School
of Medicine, 2-11-1 Kaga,
Itabashi-ku, Tokyo 173-
8606, Japan

Tada Tomohiro institute
of Gastroenterology and
proctology, Saitama, Japan
tsuoazawa244@gmail.com

Soichiro Ishihara
Tada Tomohiro institute
of Gastroenterology and
proctology, Saitama, Japan
Department of Surgical
Oncology, Graduate School
of Medicine, The University
of Tokyo, Tokyo, Japan

Mitsuhiro Fujishiro
Department of
Gastroenterology,
Graduate School of
Medicine, Nagoya
University, Nagoya, Japan

Youichi Kumagai
Department of Digestive
Tract and General
Surgery, Saitama Medical
Center, Saitama Medical
University, Saitama, Japan

Satoki Shichijo
Department of
Gastrointestinal Oncology,
Osaka International
Cancer Institute, Osaka,
Japan

Tomohiro Tada
Tada Tomohiro institute
of Gastroenterology and
proctology, Saitama, Japan
Department of Surgical
Oncology, Graduate School
of Medicine, The University
of Tokyo, Tokyo, Japan
AI medical service Inc.,
Tokyo, Japan

suggested that adenomas and SSAP have a similar relatively high risk for CRC development, whereas hyperplastic polyps rarely develop CRC.^{5,6}

Recently the American Society for Gastrointestinal Endoscopy commissioned a Preservation and Incorporation of Valuable Endoscopic Innovation (PIVI) statement to address the ‘resect and discard’ strategy.^{7,8} This strategy indicates that physicians can omit histopathological examination of resected polyps ≤ 5 mm when an optical *in vivo* diagnosis of the polyps is done with high confidence.⁸ PIVI also states that hyperplastic polyps in the rectosigmoid colon can be left in place without sampling or endoscopic resection owing to its nonmalignant nature.⁸ Therefore, accurate *in vivo* differentiation of CP leads to the reduction of unnecessary endoscopic resections, which may, in turn, decrease complications, physician burden, and medical costs.^{9,10}

Recent studies have suggested a promising role of artificial intelligence (AI) using a deep learning method in various fields, including speech recognition, visual object recognition, and object detection.^{11–13} Deep learning algorithms have shown to exceed human performance in playing certain games or in object recognition.^{14,15} In the field of medicine, previous reports have demonstrated a high potential of AI in the diagnosis of medical images, such as histology, radiography, and skin lesions.^{15–20} Thus, a computer-aided diagnosis of endoscopic images using AI has a potential to surpass the diagnostic accuracy of trained specialists and may provide more accurate results, without interobserver differences, especially between experts and nonexperts.²¹ Furthermore, it has been reported that adenoma detection rates decrease with increasing time devoted to endoscopies because of the fatigue, supporting the idea that computer-aided detection might provide more reliable results.²²

Convolutional neural networks (CNNs) are one of the most popular network architectures of deep learning for images.^{11,23} Recent studies showed a promising role for CNNs for the detection or classification of CP during colonoscopy. Those studies developed a computer-assisted diagnosis (CAD) system that can support only the detection or the classification of CP.^{24–28} However, a CAD system that can detect and simultaneously classify CP is more useful. Therefore, we developed a CAD system that can support both detection and

classification of CP during colonoscopy, and showed a high potential for future applications to real world clinical settings.

Methods

Patients

A retrospective review of clinical data from 12,895 patients who had undergone colonoscopies was performed at a single institute (Tada Tomohiro Institute of Gastroenterology and Proctology, Japan) from December 2013 to March 2017. Among them, 3021 patients were detected to have at least one polyp and underwent polypectomy. All of the specimens were examined by certified pathologists (BML Inc, Tokyo, Japan) and histologically confirmed. Patients who had histologically confirmed adenocarcinoma, adenoma, hyperplastic polyps, SSAP, juvenile polyp, Peutz-Jeghers polyp, and other polyps, such as inflammation polyps or lymphoid aggregate, were included in this study. Colonoscopy was performed using standard endoscope equipment (EVIS LUCERA and CF TYPE H260AL/I, PCF TYPE Q260AI, Q260AZI, H290I, and H290ZI; Olympus Medical Systems, Co., Ltd., Tokyo, Japan). All patient information was de-identified prior to the data analyses to maintain patient anonymity. Patients’ informed consent was exempted because of the retrospective nature of the study using completely anonymized data, and this study was approved by the Institutional Review Board of the Japan Medical Association (ID JMA-IIA00283, approved on 6 April 2017). The study protocol conforms to the ethical guidelines of the 1975 Declaration of Helsinki as reflected in a prior approval by the institution’s human research committee.

Training and validation image preparation for convolutional neural network

All of the endoscopic images of the included patients were extracted and reviewed by two trained gastroenterologists (T.O. and T.T.). Only the nonmagnified images observed using conventional white-light or narrow band imaging (NBI) mode were selected. Insufficiently insufflated colorectal images and unclear images with stool residue, halation, or bleeding were excluded from the training images. Finally, 16,418 images of 4752 histologically proven polyps from 3021 patients and 4013 images of normal colorectal

mucosa from 396 patients who had undergone colonoscopy between December 2013 and December 2016 were used to train the CNN algorithm.

For the validation set, 7077 independent images, including 1172 regions of CP from 174 patients who had undergone colonoscopy between January and March 2017 and had at least one CP, were prepared. In the validation set, even images with feces or insufficient insufflation were included to evaluate the performance of the CNN under real clinical settings. However, images from patients with inflammatory bowel disease were excluded because these may complicate the results. Images with bleeding after biopsy and images after the endoscopic treatment were also excluded. All of the polyps included in the analysis were histologically proven. Only images observed by conventional white-light or NBI mode without magnification were included, using the same criteria as those for the training image set. Detailed cohort information is shown in Table 1, and the flow chart of this study design is depicted in Figure 1a. We mainly used the NBI mode for observing the surface and vascular pattern of CP to classify the CP. Thus, the number of NBI images was relatively small compared with that of white-light images.

Algorithm for convolutional neural network

To construct an AI-based detection and diagnostic system, we utilized a deep neural network architecture called Single Shot MultiBox Detector (SSD) (<https://arxiv.org/abs/1512.02325>), without altering its algorithm.²³ SSD is a deep CNN that consists of 16 layers or more. Subsequently, a Caffe deep learning framework, originally developed at the Berkeley Vision and Learning Center, was used to train and validate the CNN. All layers of the CNN were fine-tuned using stochastic gradient descent with a global learning rate of 0.0001. Each image was resized to 300×300 pixels; the bounding box was also resized accordingly. These values were set up by trial and error to ensure all data were compatible with SSD. The authors (T.O and T.T) manually annotated all of the CP with rectangular bounding boxes and classification of polyps in the training set, and all of the images with this information were put into SSD architecture through Caffe deep learning framework (Figure 1b).

Outcome measures and statistics

First, we manually annotated all of the CP in the validation set the same as the training set ('true CP boxes'). The trained CNN also shaped the region of interests (ROIs) with rectangular bounding boxes ('CNN boxes') and output class of the CP with values ranging from 0 to 1, which showed the probability of which class the ROI belonged to. The higher the probability score, the more the CNN had confidence that the ROI included a certain class of CP.

To measure the outcome, we followed these rules: (a) when the CNN box overlapped more than 80% of the region of the true CP box, it was concluded that the CNN correctly detected the CP, and (b) when two or more CNN boxes with different classification of CP were depicted on the same region, the CNN box with highest probability score was prioritized.

We had three parameters to evaluate the performance of this CNN system in automatically detecting and classifying the images of CP: the diagnostic yields of detection and classification and the processing speed of the diagnosis. For detection performance, we analyzed (a) all images and (b) excluded CP ≥ 10 mm in size, as those CP are rarely missed.²⁹ For classification performance, we evaluated (a) all detected CP and (b) only detected CP ≤ 5 mm, to address the PIVI statement. All statistical analyses were performed using JMP Pro 10 statistical software (SAS Institute Japan, Tokyo, Japan).

Results

Association between the cut-off values of probability score and sensitivity/PPV in the validation set

To set an optimal cut-off value for the probability score to detect CP, we evaluated sensitivity and positive predictive value (PPV) by increasing the cut-off value by 0.1 from 0.1 in 10 randomly selected patients. Figure 2 shows the association between each cut-off value and sensitivity/PPV. We selected a cut-off value of 0.3 as an optimal cut-off for the probability score, in which the sensitivity and PPV were 90% and 80%, respectively. Thus, ROIs with a probability score of ≥ 0.3 were regarded as CP by the CNN.

Table 1. The information of polyps included in this study.

Training set					
Polyp type		Total polyps	WLI	NBI	Total images
Adenoma		3513 (74)	9310 (53)	2085 (73)	11,395 (56)
Hyperplastic		1058 (22)	2002 (11)	519 (18)	2521 (12)
SSAP		22 (0)	116 (1)	23 (1)	139 (1)
Cancer		68 (1)	1468 (8)	131 (5)	1599 (8)
The others		91 (2)	657 (4)	107 (4)	764 (4)
Normal		–	4013 (23)	0 (0)	4013 (20)
Total		4752 (100)	17,566 (100)	2865 (100)	20,431 (100)
Validation set					
Adenoma $n=218$	≤ 5 mm	156 (50)	639 (9)	208 (63)	847 (12)
	6–9 mm	52 (17)			
	>10 mm	10 (3)			
Hyperplastic $n=63$	≤ 5 mm	56 (18)	145 (2)	69 (21)	214 (3)
	6–9 mm	7 (2)			
	>10 mm	0 (0)			
SSAP $n=7$	≤ 5 mm	0 (0)	33 (0)	8 (2)	41 (1)
	6–9 mm	4 (1)			
	≥ 10 mm	3 (1)			
Cancer (all ≥ 10 mm) $n=4$		4 (1)	30 (0)	3 (1)	33 (0)
The others (all ≤ 5 mm) $n=17$		17 (6)	27 (0)	10 (3)	37 (1)
Normal		–	5874 (87)	31 (9)	5905 (83)
Total		309 (100)	6748 (100)	329 (100)	7077 (100)
NBI, narrow band images; SSAP, sessile serrated adenoma/polyps; WLI, white-light images. *images with multiple polyps were counted as different images.					

Diagnostic yields of detection of CP by the trained CNN

The trained CNN evaluated colonoscopy images of the validation set with a speed of 48.7 images per second, equal to a processing time of 20 ms per frame.

Figure 3a and b show the representative images in which the trained CNN properly detected and classified CP. Figure 3c shows the false-negative

(FN) case, in which the CNN missed the CP. Figure 3d shows the false-positive (FP) case, in which the CNN regarded a nonpolyp region or object as CP. Figure 3e and f represent the cases in which the CNN correctly detected the CP but misclassified them.

To evaluate the performance of the CNN in the detection of CP, we evaluated whether the CNN boxes overlapped with the true CP boxes,

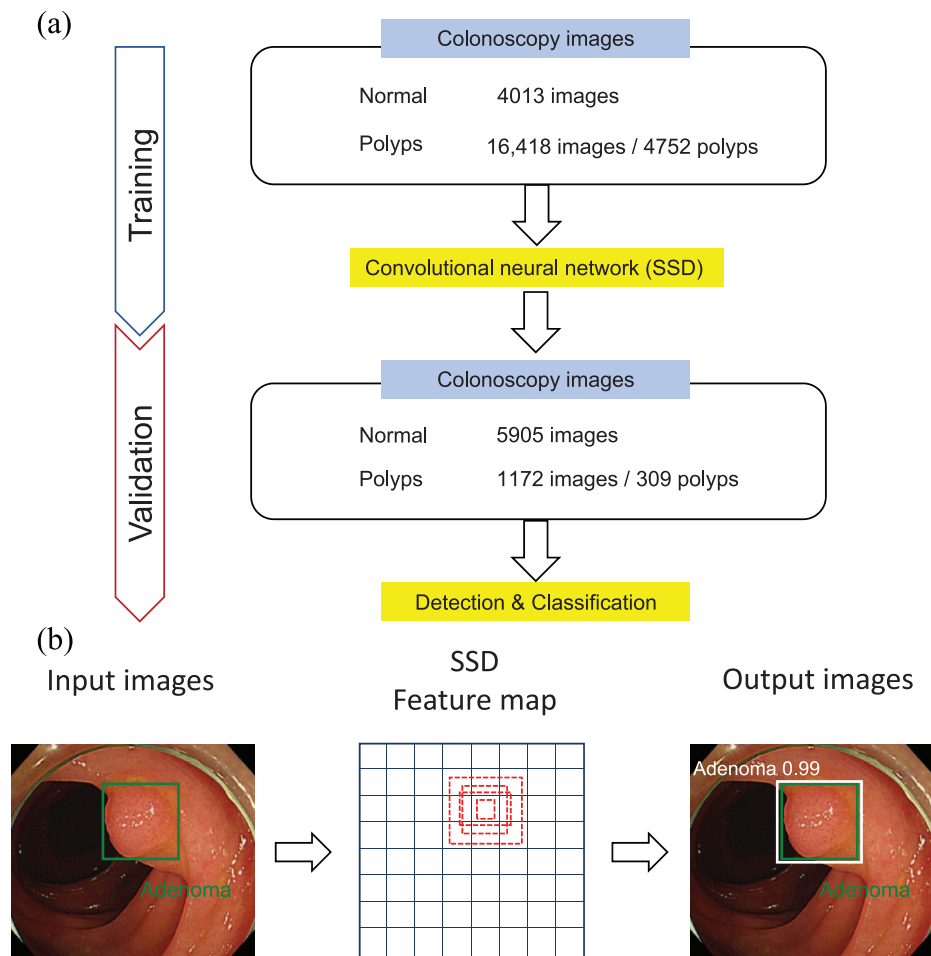


Figure 1. Study design and the convolutional neural network (CNN) used in the present study.

(a) The flow chart shows the study design. We trained the CNN using more than 20,000 colonoscopy images and validated its performance in an independent image set of 309 colorectal polyps.

(b) We used Single Shot MultiBox Detector (SSD) as a CNN, which needs an input image and bounding box (green) for each object during training. Trained CNN puts out images with bounding boxes (white) with classification of polyp and predictive score for detected object.

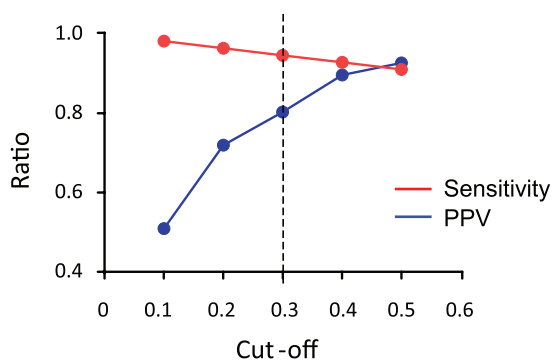


Figure 2. The association between cut-off values for the probability score and sensitivity/positive predictive values (PPV).

regardless of their classification. The CNN depicted 1246 bounding boxes (CNN boxes), and

among them, the CNN correctly detected 1073 CP out of the 1172 true CP (sensitivity 92%, PPV 86%). Although each polyp had several images, 304 CP (98%) out of 309 CP included in the validation set were detected by the trained CNN in at least one of the multiple images. By analyzing only the white-light images, the CNN demonstrated a sensitivity of 90% and PPV of 83% in the detection of CP. The CNN showed a sensitivity of 97% and PPV of 97% in only the NBI pictures, although the number of CP in NBI images was limited.

When analyzing only CP less than 10 mm in size, the CNN depicted 1143 CNN boxes, and in total, 969 boxes were overlapped with true CP boxes, showing a sensitivity of 92% and PPV of 85%, comparable with those of all CP.

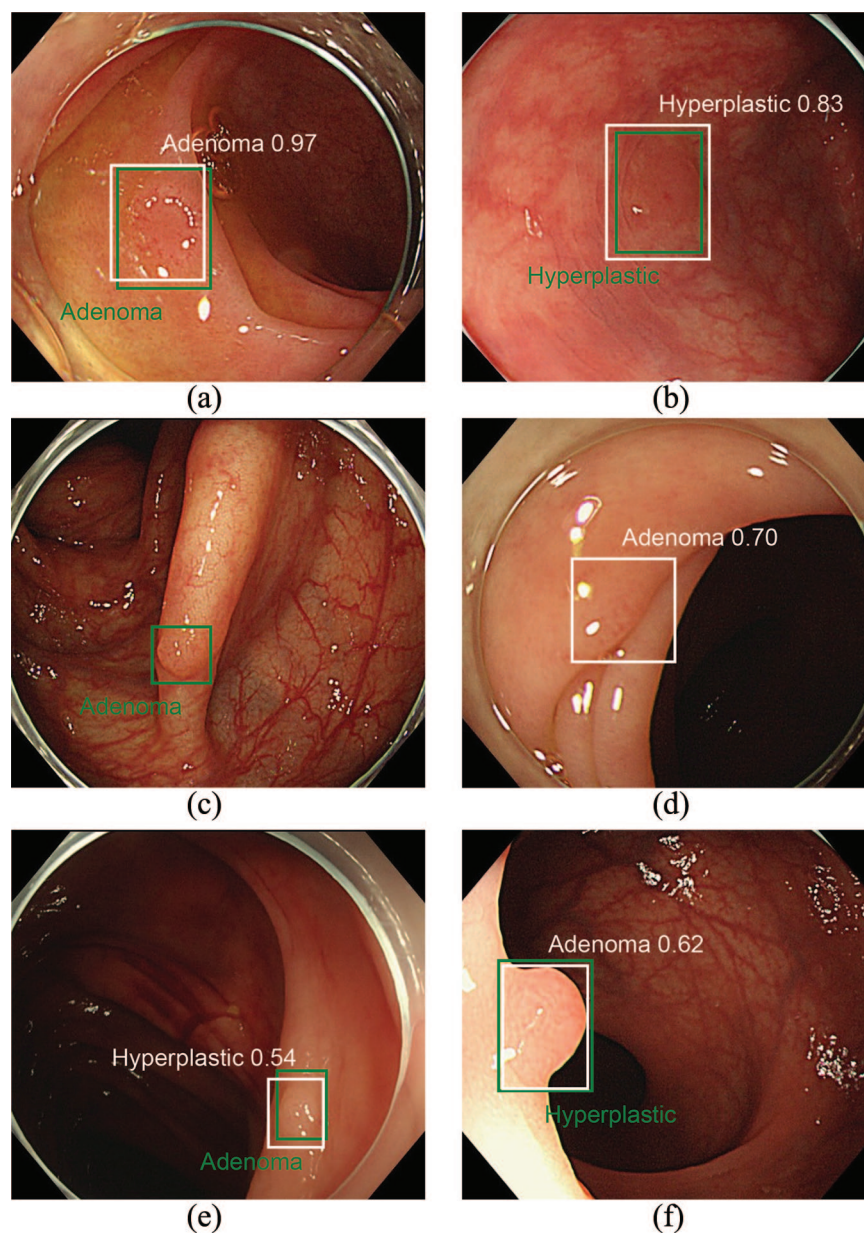


Figure 3. Representative images of colorectal polyps used in the validation set [green box: true polyp, white box: region identified as polyp by the convolutional neural network (CNN)].
 (a) Adenoma, and (b) hyperplastic polyp images correctly detected and classified by the CNN.
 (c) Adenoma image missed by the CNN (false negative image).
 (d) Normal colon fold recognized as adenoma by the CNN (false positive image).
 (e) Adenoma and (f) hyperplastic polyp images correctly detected but misclassified by the CNN.

Evaluation of the false positive region and false negative region in which the CNN identified CP
 Because it is important to evaluate why the CNN missed the CP to improve the performance of the CNN, we reviewed all of the FP and FN regions and classified them into several categories (Table 2).

A total of 99 FN polyps were classified into three categories: (a) those for which the surface texture was difficult to recognize, mainly because of small size or darkness (58%), (b) those images that were taken laterally or only partially (37%), and (c) those that were too large (5%) (Figure 4a–d).

Table 2. The distribution of the types of the images with false positive and false negative polyps.

False positive polyps (<i>n</i> = 173)		
Types	Sub-types	Numbers (%)
Normal structures	Ileocecal valve	56 (32)
	Appendiceal orifice	6 (3)
	Anus	2 (1)
Fold		56 (32)
Feces		6 (3)
True polyps?		13 (8)
The others	Halation	17 (10)
	Normal mucosa	9 (5)
	Surface haze of the camera lens	4 (2)
	Blur	2 (1)
	Scar of polypectomy	1 (1)
	Vascular dilatation	1 (1)
False negative polyps (<i>n</i> = 99)		
Hard to recognize the texture (smallness or darkness)		57 (58)
Laterality or partialness		37 (37)
Too large		5 (5)

Among 173 FP regions, 64 regions (39%) were normal structure or objects that were easy to distinguish from CP by endoscopists, most of which were ileocecal valves (*n* = 56). A total of 56 FP regions (32%) were colorectal folds, and many of them were images with insufficient insufflation. The other FPs (20%) included artificial abnormal images caused by halation (*n* = 17), haze of the lens (*n* = 4), and blur (*n* = 2), or feces, which were relatively easy to distinguish from true CP. A total of 13 regions (8%) were suspected as true CP, although we could not confirm them (Figure 4e–h).

Accuracy of classification of CP by the trained CNN

Table 3 shows the concordance between the true histology of CP and the classification of the CNN for each CP. In total, 83% of CP in conventional white-light images were correctly classified by the CNN. A total of 97% of adenomas were precisely

classified as adenoma by the CNN [PPV 86%, negative predictive value (NPV) 85%] when analyzed without cancers in conventional white-light images only, although only 47% of hyperplastic polyps were correctly identified as hyperplastic polyps (PPV 64%, NPV 90%), and many SSAPs were misclassified as adenoma (26%) or hyperplastic polyps (52%). Similarly, 81% of CP in NBI images were correctly classified, and the sensitivity to classify adenomas from the other polyps was 97% (PPV 83%, NPV 91%) in NBI images, although the number of NBI images was limited.

We also analyzed the performance of the CNN in the classification of CP \leq 5 mm in size. The CNN correctly classified 348 (98%) out of 356 adenoma images (PPV 85%, NPV of 88%) in conventional white-light images, although for hyperplastic polyps, the classification performance was modest (sensitivity 50%, PPV 77%, NPV 88%). In NBI mode, the CNN accurately classified 138 (97%) out of 142 adenoma images

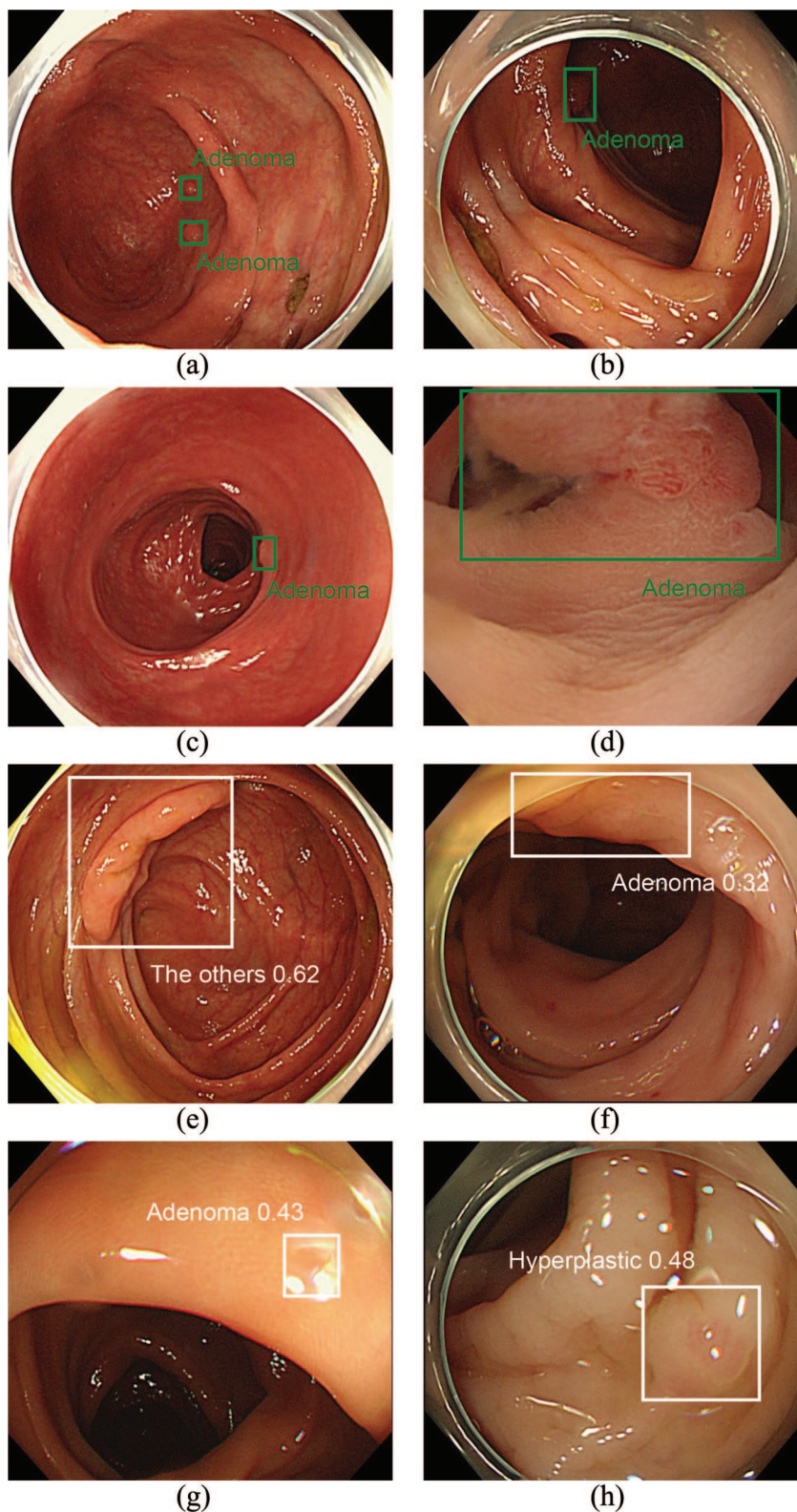


Figure 4. Representative images of false positive and false negative polyps used in the validation set (green box: true polyp, white box: region identified as polyp by the convolutional neural network). (a)–(d) Images of false negative polyps were classified into three types: (1) hard to recognize the texture, because of the small size of the polyps (a), and darkness (b); (2) polyp images were taken partially or laterally (c); (3) polyp images were relatively large (d). (f)–(h) Images of false positive polyps were classified into four types: (1) normal structure such as ileocecal valve (e); (2) normal colorectal fold (f); (3) artificial images such as halation (g); (4) suspected true polyps (h).

Table 3. The distribution of the types of polyps classified by the CNN.

		White light images (n = 783)				
		The CNN classification (%)				
		Adenoma	Hyperplastic	SSAP	Cancer	The others
True histology	Adenoma (n = 582)	562 (97)	14 (2)	0 (0)	4 (1)	2 (0)
	Hyperplastic (n = 125)	64 (51)	59 (47)	0 (0)	0 (0)	2 (2)
	SSAP (n = 23)	6 (26)	12 (52)	5 (22)	0 (0)	0 (0)
	Cancer (n = 29)	6 (21)	0 (0)	0 (0)	23 (79)	0 (0)
	The others (n = 24)	14 (58)	7 (29)	0 (0)	0 (0)	3 (13)
		Narrow band images (n = 290)				
		The CNN classification (%)				
		Adenoma	Hyperplastic	SSAP	Cancer	The others
True histology	Adenoma (n = 203)	197 (97)	5 (2)	0 (0)	1 (0)	0 (0)
	Hyperplastic (n = 68)	31 (46)	37 (54)	0 (0)	0 (0)	0 (0)
	SSAP (n = 6)	2 (33)	4 (67)	0 (0)	0 (0)	0 (0)
	Cancer (n = 3)	3 (100)	0 (0)	0 (0)	0 (0)	0 (0)
	The others (n = 10)	3 (30)	7 (70)	0 (0)	0 (0)	0 (0)

CNN, convolutional neural network; SSAP, sessile serrated adenoma/polyp.

(PPV 84%, NPV of 88%), although the number of NBI images was limited (Table 4). These results show that the performance of the CNN in the classification of CP was comparable regardless of polyp size.

Discussions

In the present study, we, for the first time to our knowledge, demonstrated the CNN-based detection and classification of CP using a large number of image sets. Our trained CNN effectively detected CP with considerable accuracy and surprising speed, even when the CP were small, which may help reduce overlooked CP if applied during colonoscopy. Furthermore, the CNN classified and detected CP with considerable performance, which may help reduce unnecessary treatment benefitting both patients and physicians.

We applied SSD, a meta-architecture and feature extractor, to develop a deep-learning-based system to detect and classify polyps.^{23,30} The SSD

performs object recognition by using a feed-forward convolutional network that produces a fixed-size collection of bounding boxes and scores for the presence of an object class in each box. This SSD can deal with objects of various sizes by combining predictions from multiple feature maps with different resolutions. Furthermore, it encapsulates the process into a single network, thus saving computational time. Currently, there are several high-performance meta-architectures that are suitable for this purpose. Fuentes and colleagues recently reported a deep-learning-based detector to recognize tomato plant diseases and pests combining three detectors, including SSD, and reported a high degree of accuracy.³⁰ Therefore, by increasing the training images and by modifying the architecture itself, the accuracy of the CNN may be improved although our CNN has demonstrated a considerably good performance already.²³

The SSD algorithm enabled the CNN to not only detect CP, but also to classify the CP. This

Table 4. The distribution of the types of diminutive polyps classified by the CNN.

		White light images (n = 480)		
		The CNN classification (%)		
		Adenoma	Hyperplastic	The others
True histology	Adenoma (n = 356)	348 (98)	8 (2)	0
	Hyperplastic (n = 100)	49 (49)	50 (50)	1 (1)
	The others (n = 24)	14 (58)	7 (29)	3 (13)
		Narrow band images (n = 198)		
		The CNN classification (%)		
		Adenoma	Hyperplastic	The others
True histology	Adenoma (n = 142)	138 (97)	4 (3)	0 (0)
	Hyperplastic (n = 46)	24 (52)	22 (48)	0 (0)
	The others (n = 10)	3 (30)	7 (70)	0 (0)

CNN, convolutional neural network.

system is more useful than the CAD systems that can perform either detection or diagnosis of CP to achieve the ‘resect and discard strategy’. The trained CNN classified adenomas, which are subject to endoscopic resection, with a sensitivity of 97% and an accuracy of 87% (analyzed excluding cancers) in conventional white-light images. It has been reported that white-light colonoscopy has only a limited accuracy of 59–84% in differentiating nonneoplastic polyps from neoplastic polyps.^{7,31,32} Furthermore, according to PIVI statements, ‘In order for a technology to be used to guide the decision to leave suspected rectosigmoid hyperplastic polyps ≤ 5 mm in size in place, the technology should provide $\geq 90\%$ NPV for adenomatous histology.’²⁸ Our trained CNN classified adenomas with NPVs of 85% and 91% by white-light image and by NBI, respectively, and these results were comparable when analyzing only small CP (≤ 5 mm in size). The CNN also provides completely objective classification with a probability score that is an important issue in the decision making of the ‘resect and discard’ or ‘leave rectosigmoid colon hyperplastic polyps *in situ*’ policies for CP. Therefore, the CNN-based CP diagnostic system is a highly promising technology for ‘optical biopsy’ during colonoscopy.

Byrne and colleagues recently reported an AI-based model for real-time differentiation of adenomatous

and hyperplastic diminutive polyps during standard colonoscopy.³³ In their study, the authors trained their CNN using colonoscopy video in NBI mode only, and developed an AI system that can effectively distinguish surface patterns of polyps under NBI. We mainly used white-light images during colonoscopy and utilized the NBI mode to evaluate the histology for a limited number of CP. Therefore, in the present study, the number of the training images in NBI mode were not sufficient to make the CNN learn the surface pattern of each polyp. However, our trained CNN distinguished the histology of CP under NBI mode better than white-light mode. Thus, for a future study, it will be useful to know that if by learning more NBI images, the CNN will improve detection performance or classification ability. Furthermore, it is also fascinating to evaluate the performance of CNNs that have learned new imaging technologies, such as blue laser and autofluorescence, for the detection or classification of CP.^{34,35}

We acknowledge several limitations of the present study. First, this is a retrospective study in a single institute, thus, external validation and a prospective study is necessary to evaluate the performance of our CNN. In particular, it is important to evaluate whether the CNN really supports physicians’ performance of colonoscopy in terms of detection rate and classification

accuracy of CP. In this regard, recently Wang and colleagues conducted a double-blind randomized study and showed that their deep-learning computer-aided system increased adenoma detection rate.³⁶ Second, to improve the accuracy of the CNN, it is important to use a sufficient number of training images. In this regard, we used more than 15,000 histologically proven various types of CP images as the training set. However, the performance to identify adenoma through normal white light was not satisfactory for the PIVI statement, and half of the hyperplastic polyps were regarded as adenoma by the trained CNN. The performance of the CNN to classify CP may be underestimated because these analyses included CP images that were not taken close enough to observe the surface pattern, however, these results show that we have still room to further improve our CNN by increasing the amount of training images, including enhanced images, or modifying the CNN architecture. Furthermore, training images used in this study have selection biases, as many training images were ‘clear’ and ‘right size’ images that are among the causes of overlooking small CP, while there were more unclear images in the validation set. Therefore, we are collecting those images of polyps as well to make a much more powerful CNN that can be applied to real clinical settings. Finally, the present study was conducted in only still images. However, the processing time of our CNN is fast enough to be applied to real-time video images that require processing time of less than 30 ms per frame, and we are now conducting a prospective study using present CNN during colonoscopy with a real-time manner.

In conclusion, we developed and evaluated the CNN-based detector and classifier of CP using large numbers of colonoscopy images. Our trained CNN showed a robust performance to detect and classify CP and may be used as a CNN-based colonoscopy supporting system.

Author contributions

T.O. was involved in study concept and design, analysis and interpretation of data and drafting of the manuscript. M.F., Y.K., and S.S. were involved in critical revision of the manuscript for important intellectual content and material support. T.T. was involved in study concept, acquisition of data, analysis, interpretation of data, and study supervision.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research is financially supported by the Japanese Grans-in-Aid for Scientific Research (grant number 19K16810) from Japan Society for the promotion of Science.

Conflict of interest statement

T.T is an employee of AI Medical Service Inc.

ORCID iD

Tsuyoshi Ozawa  <https://orcid.org/0000-0002-0701-7978>


References

1. Manceau G and Panis Y. Laparoscopic colorectal surgery: why, when, how? *Updates Surg* 2016; 68: 3–5.
2. Siegel RL, Miller KD and Jemal A. Cancer statistics, 2016. *CA Cancer J Clin* 2016; 66: 7–30.
3. Strum WB. Colorectal adenomas. *N Engl J Med* 2016; 374: 1065–1075.
4. US Preventive Services Task Force, Bibbins-Domingo K, Grossman DC, *et al.* Screening for colorectal cancer: US preventive services task force recommendation statement. *JAMA* 2016; 315: 2564–2575.
5. JE IJ, Bastiaansen BA, van Leerdam ME, *et al.* Development and validation of the WASP classification system for optical diagnosis of adenomas, hyperplastic polyps and sessile serrated adenomas/polyps. *Gut* 2016; 65: 963–970.
6. JEG IJ, Bevan R, Senore C, *et al.* Detection rate of serrated polyps and serrated polyposis syndrome in colorectal cancer screening cohorts: a European overview. *Gut* 2017; 66: 1225–1232.
7. Ignjatovic A, East JE, Suzuki N, *et al.* Optical diagnosis of small colorectal polyps at routine colonoscopy (Detect InSpect ChAracterise Resect and Discard; DISCARD trial): a prospective cohort study. *Lancet Oncol* 2009; 10: 1171–1178.
8. Rex DK, Kahi C, O’Brien M, *et al.* The American society for gastrointestinal endoscopy PIVI (preservation and incorporation of valuable endoscopic innovations) on real-time endoscopic assessment of the histology of diminutive colorectal polyps. *Gastrointest Endosc* 2011; 73: 419–422.

9. Sikka S, Ringold DA, Jonnalagadda S, *et al.* Comparison of white light and narrow band high definition images in predicting colon polyp histology, using standard colonoscopes without optical magnification. *Endoscopy* 2008; 40: 818–822.
10. Hassan C, Pickhardt PJ and Rex DK. A resect and discard strategy would improve cost-effectiveness of colorectal cancer screening. *Clin Gastroenterol Hepatol* 2010; 8: 865–869, 869. e861–e863.
11. LeCun Y, Bengio Y and Hinton G. Deep learning. *Nature* 2015; 521: 436–444.
12. Zhang R, Zheng Y, Mak TW, *et al.* Automatic detection and classification of colorectal polyps by transferring low-level CNN features from nonmedical domain. *IEEE J Biomed Health Inform* 2017; 21: 41–47.
13. Sainath TN, Kingsbury B, Saon G, *et al.* Deep convolutional neural networks for large-scale speech tasks. *Neural Netw* 2015; 64: 39–48.
14. Silver D, Schrittwieser J, Simonyan K, *et al.* Mastering the game of Go without human knowledge. *Nature* 2017; 550: 354–359.
15. Mnih V, Kavukcuoglu K, Silver D, *et al.* Human-level control through deep reinforcement learning. *Nature* 2015; 518: 529–533.
16. Lakhani P and Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology* 2017; 284: 574–582.
17. Esteva A, Kuprel B, Novoa RA, *et al.* Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017; 542: 115–118.
18. Xu Y, Jia Z, Wang LB, *et al.* Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features. *BMC Bioinformatics* 2017; 18: 281.
19. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, *et al.* Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* 2017; 318: 2199–2210.
20. Ciompi F, Chung K, van Riel SJ, *et al.* Towards automatic pulmonary nodule management in lung cancer screening with deep learning. *Sci Rep* 2017; 7: 46479.
21. Postgate A, Tekkis P, Fitzpatrick A, *et al.* The impact of experience on polyp detection and sizing accuracy at capsule endoscopy: implications for training from an animal model study. *Endoscopy* 2008; 40: 496–501.
22. Almadi MA, Sewitch M, Barkun AN, *et al.* Adenoma detection rates decline with increasing procedural hours in an endoscopist's workload. *Can J Gastroenterol Hepatol* 2015; 29: 304–308.
23. Liu W, Anguelov D, Erhan D, *et al.* SSD: single shot multibox detector. Cham: Springer International Publishing, 2016, pp. 21–37.
24. Chen PJ, Lin MC, Lai MJ, *et al.* Accurate classification of diminutive colorectal polyps using computer-aided analysis. *Gastroenterology* 2018; 154: 568–575.
25. Misawa M, Kudo SE, Mori Y, *et al.* Artificial intelligence-assisted polyp detection for colonoscopy: initial experience. *Gastroenterology* 2018; 154: 2027–2029. e2023.
26. Urban G, Tripathi P, Alkayali T, *et al.* Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy. *Gastroenterology* 2018; 155: 1069–1078. e1068.
27. Klare P, Sander C, Prinzen M, *et al.* Automated polyp detection in the colorectum: a prospective study (with videos). *Gastrointest Endosc* 2019; 89: 576–582. e571.
28. Wang P, Berzin TM, Glissen Brown JR, *et al.* Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. *Gut* 2019; 68: 1813–1819.
29. van Rijn JC, Reitsma JB, Stoker J, *et al.* Polyp miss rate determined by tandem colonoscopy: a systematic review. *Am J Gastroenterol* 2006; 101: 343–350.
30. Fuentes A, Yoon S, Kim SC, *et al.* A Robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. *Sensors (Basel)* 2017; 17: pii: E2022.
31. Machida H, Sano Y, Hamamoto Y, *et al.* Narrow-band imaging in the diagnosis of colorectal mucosal lesions: a pilot study. *Endoscopy* 2004; 36: 1094–1098.
32. Su MY, Hsu CM, Ho YP, *et al.* Comparative study of conventional colonoscopy, chromoendoscopy, and narrow-band imaging systems in differential diagnosis of neoplastic and nonneoplastic colonic polyps. *Am J Gastroenterol* 2006; 101: 2711–2716.
33. Byrne MF, Shahidi N and Rex DK. Will computer-aided detection and diagnosis

- revolutionize colonoscopy? *Gastroenterology* 2017; 153: 1460–1464.e1461.
34. Takeuchi Y, Inoue T, Hanaoka N, *et al.* Autofluorescence imaging with a transparent hood for detection of colorectal neoplasms: a prospective, randomized trial. *Gastrointest Endosc* 2010; 72: 1006–1013.
35. Yoshida N, Yagi N, Inada Y, *et al.* Ability of a novel blue laser imaging system for the diagnosis of colorectal polyps. *Dig Endosc* 2014; 26: 250–258.
36. Wang P, Liu X, Berzin TM, *et al.* Effect of a deep-learning computer-aided detection system on adenoma detection during colonoscopy (CADe-DB trial): a double-blind randomised study. *Lancet Gastroenterol Hepatol*. Epub ahead of print 22 January 2020. DOI: 10.1016/S2468-1253(19)30411-X.

Visit SAGE journals online
[journals.sagepub.com/
home/tag](https://journals.sagepub.com/home/tag)

 SAGE journals