Automatic detection and classification of protruding lesions in wireless capsule endoscopy images based on a deep convolutional neural network CME

Hiroaki Saito, MD,[1] Tomonori Aoki, MD, PhD,[2] Kazuharu Aoyama, MMath,[3] Yusuke Kato, BSc,[3] Akiyoshi Tsuboi, MD,[4] Atsuo Yamada, MD, PhD,[2] Mitsuhiro Fujishiro, MD, PhD,[5] Shiro Oka, MD, PhD,[4] Soichiro Ishihara, MD, PhD,[6,7] Tomoki Matsuda, MD, PhD,[1] Masato Nakahori, MD, PhD,[1] Shinji Tanaka, MD, PhD,[4] Kazuhiko Koike, MD, PhD,[2] Tomohiro Tada, MD, PhD[3,6,7]

Tokyo, Japan

Background and Aims: Protruding lesions of the small bowel vary in wireless capsule endoscopy (WCE) images, and their automatic detection may be difficult. We aimed to develop and test a deep learning–based system to automatically detect protruding lesions of various types in WCE images.

Methods: We trained a deep convolutional neural network (CNN), using 30,584 WCE images of protruding lesions from 292 patients. We evaluated CNN performance by calculating the area under the receiver operating characteristic curve (AUC), sensitivity, and specificity, using an independent set of 17,507 test images from 93 patients, including 7507 images of protruding lesions from 73 patients.

Results: The developed CNN analyzed 17,507 images in 530.462 seconds. The AUC for detection of protruding lesions was 0.911 (95% confidence interval [CI], 0.9069–0.9155). The sensitivity and specificity of the CNN were 90.7% (95% CI, 90.0%–91.4%) and 79.8% (95% CI, 79.0%–80.6%), respectively, at the optimal cut-off value of 0.317 for probability score. In a subgroup analysis of the category of protruding lesions, the sensitivities were 86.5%, 92.0%, 95.8%, 77.0%, and 94.4% for the detection of polyps, nodules, epithelial tumors, submucosal tumors, and venous structures, respectively. In individual patient analyses (n = 73), the detection rate of protruding lesions was 98.6%.

Conclusion: We developed and tested a new computer-aided system based on a CNN to automatically detect various protruding lesions in WCE images. Patient-level analyses with larger cohorts and efforts to achieve better diagnostic performance are necessary in further studies. (Gastrointest Endosc 2020;92:144-51.)

Wireless capsule endoscopy (WCE) is a useful examination method for detecting and diagnosing small-bowel diseases.[1,2] WCE can survey the entire digestive tract from the esophagus to the colon with approximately 60,000 images. Of this large number of images, abnormal findings may appear in only 1 or 2 frames, so that close observation by physicians for approximately 1 to 2 hours is required for a correct

Abbreviations: AI, artificial intelligence; AUC, area under the curve; CEST, capsule endoscopy structured terminology; CI, confidence interval; CNN, convolutional neural network; FAP, familial adenomatous polyposis; IoU, intersection over union; PS, probability score; ROC, receiver operating characteristic; SMT, submucosal tumors; SSD, single shot multibox detector; WCE, wireless capsule endoscopy.

DISCLOSURE: *All authors disclosed no financial relationships.*

See CME section; p. 198.

Copyright © 2020 by the American Society for Gastrointestinal Endoscopy
0016-5107/$36.00
https://doi.org/10.1016/j.gie.2020.01.054

Received October 31, 2019. Accepted January 31, 2020.

Current affiliations: Department of Gastroenterology, Sendai Kousei Hospital, Miyagi (1), Department of Gastroenterology, Graduate School of Medicine, The University of Tokyo, Tokyo (2), AI Medical Service Inc., Tokyo (3), Department of Endoscopy, Hiroshima University Hospital, Hiroshima (4), Department of Gastroenterology and Hepatology, Nagoya University Graduate School of Medicine, Aichi (5), Department of Surgical Oncology, Graduate School of Medicine, The University of Tokyo, Tokyo (6), Tada Tomohiro Institute of Gastroenterology and Proctology, Saitama (7), Japan.

Reprint requests: Hiroaki Saito, MD, Department of Gastroenterology; Sendai Kousei Hospital, 4-15, Hirose-cho, Aoba-ku, Sendai, Miyagi 9800873, Japan. E-mail: h.saito0515@gmail.com.

If you would like to chat with an author of this article, you may contact Dr Saito at h.saito0515@gmail.com.

diagnosis.[3,4] The interpretation of WCE images is a great burden on the gastroenterologist.[4,5]

WCE can detect several types of protruding lesions in the small bowel. Their morphologic characteristics vary, from polyps, nodules, masses, and tumors, to venous structures.[6] The causes underlying these lesions also vary, including neuroendocrine tumors, adenocarcinomas, familial adenomatous polyposis, Peutz-Jeghers syndrome, follicular lymphomas, and GI stromal tumors. Because these lesions require early diagnosis and treatment, oversights during capsule endoscopy readings should be avoided.

Recently, image analysis using artificial intelligence (AI) in many medical fields, including endoscopy, has been developed.[7,8] In particular, deep learning with the use of convolutional neural networks (CNNs), which learn the characteristics of images based on accumulated images, has attracted attention as being useful for the analysis of medical images and for image-based diagnosis. CNNs, which can automatically and rapidly process large numbers of images, are expected to be applicable to capsule endoscopy. Although CNNs can achieve acceptable results in the field of WCE of the small intestine,[9-11] there are still many unresolved issues with the application of CNNs for the detection of lesions in the small intestine. As concerns protruding lesions, analyses based on the detailed classification of lesions according to morphology remain scarce. The number of protruding lesions, and the number of morphologic or etiologic types of protruding lesions used for analysis, were small in previous studies.

In the present study, we trained a CNN on the basis of WCE images of small bowels collected retrospectively from multiple institutions and constructed a model to automatically detect protruding lesions.

METHODS

Preparation of CNN training images

WCE images were collected retrospectively from 3 affiliated institutions: Sendai Kousei Hospital, the University of Tokyo, and Hiroshima University Hospital, from October 2009 to May 2018. During that period, WCEs were performed with the Pillcam SB2 or SB3 (Given Imaging, Yoqneam, Israel). All cases that included images with protruding lesions in that period were comprehensively extracted. The cause of protruding lesions was clinically diagnosed based on findings in WCE and other subsequent examinations at each institution. We collected 30,584 images of protruding lesions from 292 patients and used them as a training image dataset (Fig. 1).

Images used for testing the CNN

A total of 17,507 images from 93 patients were used for testing the CNN, including 10,000 images without lesions and 7507 images with protruding lesions (Fig. 1). These images were different from the images used for training the CNN.

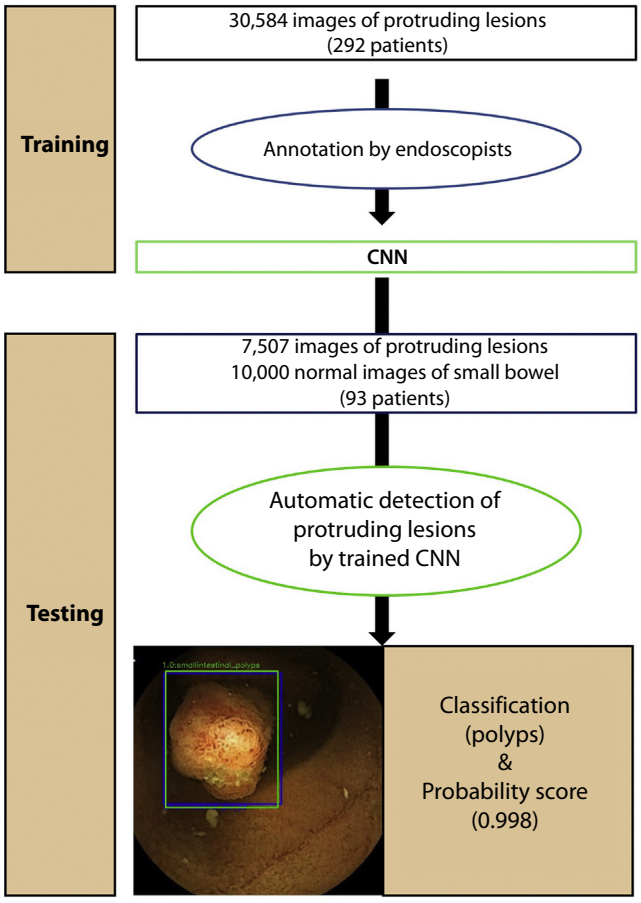


Figure 1. Flow chart of the study design. *CNN,* Convolutional neural network

Definition and classification of protruding lesions

Protruding lesions were morphologically classified into 5 categories as polyps, nodules, epithelial tumors, submucosal tumors (SMT), and venous structures, based on a capsule endoscopy structured terminology (CEST) classification.[6] In these categories, we classified masses and tumors based on CEST definitions into epithelial tumors and SMT.

CNN algorithms

To construct an AI-based diagnostic system, we used a deep neural network architecture called the single shot multiBox detector (SSD)[12] without altering its underlying algorithm. SSD is a deep CNN that consists of 16 or more layers. All regions showing protruding lesions in the training set of images were manually annotated for this study by 6 expert endoscopists (H.S, M.N, A.Y, T.A., A.T., and S.O.) with rectangular bounding boxes. The annotation was performed by each endoscopist separately, and consensus was determined subsequently. These images were fed into the SSD architecture through the Caffe deep learning framework.[13] This framework, which was used to train and test the CNN in this study, is one of the most popular and widely used

frameworks and was originally developed at the Berkeley Vision and Learning Center. The CNN was taught to recognize the areas within the bounding boxes as representing protruding regions and the other areas as representing background. All layers of the CNN were fine-tuned by the use of 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. The rotation of images was added for data augmentation. To acquire a high-performance CNN model, the appropriate values for hyperparameters, such as learning rate, were set up by repeated trial and error. For tuning model hyperparameters, 20% of a training dataset was randomly chosen as a validation set for each type of protruding lesion. For calculation, we used an Intel Core i7-7700 (intel Corp, Santa Clara, Calif, USA) as the central processing unit and a GeForce GTX 1070 (Nvidia, Santa Clara, Calif, USA) as the graphics processing unit.

Outcome measures and statistics

The primary outcome was the detection of protruding lesions by CNN based on the area under the receiver operating characteristic curve (ROC-AUC), sensitivity, and specificity. The sensitivity of detection by the CNN was examined for each category of protruding lesions as a subanalysis. Accuracy of classification of protruding lesions by the CNN was not considered in the primary outcome. Testing was performed on a single-image basis. First, 6 endoscopists (H.S., M.N., A.Y., T.A., A.T., and S.O.) manually annotated the areas of all protruding lesions in the test set with blue square boxes ("true" boxes). The constructed CNN enclosed the protruding lesions by green square boxes (CNN boxes), with a category name of protruding lesions, and a probability score (PS) ranging from 0 to 1, which represented CNN's confidence. We evaluated the ability of the CNN to determine whether or not each image contained a protruding lesion.

We evaluated CNN boxes in descending order of PS for each image and evaluated the overlap between CNN box and "true" box using intersection over union (IoU), which is an evaluation method to measure the accuracy of an object detector; an IoU was calculated by dividing the area of overlap of 2 boxes by their combined area. The higher the IoU, the greater the overlap and the better the result. The additional explanation of IoU is in the legend for Supplementary Figure 1A–C (available online at www.giejournal.org). We reviewed boxes in all images and defined the CNN box with $\geq$0.05 IoU as overlapped.

The following images were defined as correct answers by the CNN: (1) In terms of an image including 1 protruding lesion, a CNN box overlapped the "true" box. (2) In terms of an image including a plurality of "true" boxes, a CNN box overlapped with at least 1 "true" box. (3) In terms of an image without protruding lesions, the PS of the CNN boxes in the image did not exceed the cut-off value. These tasks were performed on all images by 3 endoscopists (H.S., T.A., and A.Y.).

The ROC was plotted by varying the threshold of the PS, and the AUC was calculated to assess the degree of discrimination. The sensitivity, specificity, and accuracy of the CNN in detecting protruding lesions were calculated using various cut-off values for the PS, including the score according to the Youden index. The Youden index is one of the standard methods for determining the most ideal cutoff values using sensitivity and specificity.[14]

The secondary outcomes were (1) the concordance between CNN and expert endoscopists for classification of protruding lesions into 5 categories and (2) the detection of protruding lesions in individual patient analyses. In individual patient analyses aimed at determining the detection rate of protruding lesions, detection by the CNN was defined as correct when the CNN detected at least 1 image from multiple images of the same patient. In addition, we reevaluated the 10,000 clinically normal images in the test set after using the CNN. CNN boxes in normal images that appeared to be true lesions were extracted; these lesions might have been overlooked by physicians. This was based on the consensus of 3 expert endoscopists (H.S., A.T., and A.Y.).

The data were analyzed with STATA (version 15; StataCorp, College Station, Tex, USA). This study was approved by the Ethics Committee of the Japan Medical Association (ID JMA-IIA00283), Sendai Kosei Hospital (No. 30-5), the University of Tokyo (No. 11931), and Hiroshima University Hospital (No. E-1246). This was a retrospective study using anonymized data, carried out by the opt-out method of our hospital website or outpatient department.

RESULTS

CNN capability for detecting protruding lesions

Details of the cases used for training and tests are shown in Table 1. The test set consisted of 7507 images with protruding lesions from 73 patients (men, 65.8%; mean age, 60.1 years; standard deviation, 18.7 years) and 10,000 images without lesions from 20 patients (men, 60.0%; mean age, 51.9 years; standard deviation, 11.4 years). In the testing dataset, the most common cause of protruding lesions was malignant lymphoma (25%), and the most common cause of polyps was hamartomatous polyp (15%). The constructed CNN analyzed the images in 530.462 seconds, with an average speed of 0.0303 seconds per image. The AUC of the CNN used to detect protruding lesions was 0.911 (95% CI, 0.9069–0.9155) (Supplementary Fig. 2, available online at www.giejournal.org). Table 2 shows each sensitivity and specificity by increasing the cut-off value for the probability score by 0.1 from 0.1 to 0.9. According to the Youden index, the optimal cut-off value for the PS was 0.317; thus, the regions with a PS of 0.317 or more were recognized as protruding lesions detected by the CNN. When that cut-off value was used, the

TABLE 1. Patient characteristics of training and test datasets

Characteristics, n (%)	Training dataset	Test dataset	
		Protruding lesions	Normal
No. of images	30,584	7507	10,000
Polyps	10,704 (35.0)	3522 (46.9)	-
Nodules	11,097 (36.3)	1932 (25.7)	-
Epithelial tumors	6515 (21.3)	1462 (19.5)	-
Submucosal tumors	1875 (6.1)	339 (4.5)	-
Venous structures	393 (1.3)	252 (3.4)	-
No. of patients	292	73	20
Mean age, years ± SD	61.1 ± 17.0	60.1 ± 18.7	51.9 ± 11.4
Sex, male	184 (63.0)	48 (65.8)	12 (60.0)
Protruding lesions			
Polyps	78 (26.7)	30 (41.1)	-
Nodules	78 (26.7)	14 (19.2)	-
Epithelial tumors	36 (12.3)	14 (19.2)	-
Submucosal tumors	66 (22.6)	11 (15.1)	-
Venous structures	34 (11.6)	4 (5.5)	-
No. of lesions per patient			
Single	133 (45.6)	37 (50.7)	-
Multiple	159 (54.4)	36 (49.3)	-
Indication for capsule endoscopy			
Obscure GI bleeding	93 (31.8)	27 (37.0)	
Lymphoma	76 (26.0)	13 (17.8)	
Polyposis	32 (11.0)	12 (16.4)	
Abnormal small-bowel imaging in other modalities	63 (21.6)	11 (15.1)	
Screening	19 (6.5)	9 (12.3)	
Abdominal pain	8 (2.7)	1 (1.4)	
Diarrhea	1 (0.3)	0 (0)	
Diagnosis*			
Hamartomatous polyp	31 (10.6)	11 (15.1)	-
Adenomatous polyp	23 (7.9)	5 (6.8)	-
Inflammatory polyp	3 (1.0)	4 (5.5)	-
Malignant lymphoma	95 (32.5)	18 (24.7)	-
Cancer	12 (4.1)	6 (8.2)	-
Lipomatosis	15 (5.1)	3 (4.1)	-
GIST	9 (3.1)	0 (0)	-
Carcinoid	3 (1.0)	1 (1.4)	-
Hemangioma	15 (5.1)	2 (2.7)	-
Varices	8 (2.7)	1 (1.4)	-
Others	33 (11.3)	10 (13.7)	-
Unknown	45 (15.4)	12 (16.4)	-

SD, Standard deviation; *GIST,* GI stromal tumor.
*The diagnosis was based on other endoscopic and surgical findings and pathologic examinations after capsule endoscopy.

sensitivity and specificity of the CNN were 90.7% (95% CI, 90.0%–91.4%) and 79.8% (95% CI, 79.0%–80.6%), respectively (Table 3). In a subanalysis of the categories of protruding lesions, the values for CNN sensitivity were 86.5%, 92.0%, 95.8%, 77.0%, and 94.4% for the detection of polyps, nodules, epithelial tumors, SMTs, and venous structures, respectively. Figure 2 shows representative regions correctly detected and classified by the CNN in 5 categories.

TABLE 2. Ability to detect protruding lesions for each cut-off value

Cut-off value (probability score)	Sensitivity, %	Specificity, %
0.1	99.7	9.9
0.2	97.9	60.6
0.3	92.3	77.9
0.317*	90.7	79.8
0.4	79.7	86.2
0.5	65.5	90.8
0.6	52.9	93.7
0.7	41.6	96.0
0.8	29.5	97.7
0.9	16.3	99.0

*Calculated according to the Youden index.

TABLE 3. Detection of protruding lesions by the convolutional neural network (CNN)

	Experts' diagnoses		
The CNN's diagnosis	Protruding lesions	Normal mucosa	Total
Protruding lesions	6810 (90.7)	2019 (20.2)	8678
Normal mucosa	697 (9.3)	7981 (79.8)	8829
Total	7507	10,000	17,507
	Sensitivity: 90.7%	Specificity: 79.8%	

According to the Youden index, regions with a probability score of ≥0.317 were recognized as protruding lesions by the CNN
CNN, Convolutional neural network.

In individual patient analyses, the detection rate of the protruding lesions was 98.6% (72/73). By category of protruding lesion, the detection rates per patient for polyps, nodules, epithelial tumors, SMTs, and venous structures were 96.7% (29/30), 100% (14/14), 100% (14/14), 100% (11/11), and 100% (4/4), respectively. In 1 patient with familial adenomatous polyposis (FAP), not all 3 images of polyps could be detected by the CNN (Fig. 3).

As shown in Table 4, false-negative images by the CNN at the cut-off value of 0.317 for the PS (n = 697) were classified into 4 categories based on the cause of the false-negative read: similar color to the surrounding healthy mucosa (n = 277), partialness or laterality (n = 146), small size (n = 96), and poor demarcation (n = 178). By contrast, false-positive images (n = 2019) were classified into 4 categories based on the reason behind the false-positive read: normal mucosa (n = 1844), foam (n = 132), debris (n = 37), and vascular dilatation (n = 6). Among the false-positive images for which the CNN provided boxes, 2 images were suggested to include true protruding lesions by the endoscopists (Fig. 4).

CNN capability for classifying protruding lesions

The labeling of protruding lesions by the CNN and expert endoscopists is presented in Table 5. The rates of concordance of the labeling by the CNN and the endoscopists for polyps, nodules, epithelial tumors, SMTs, and venous structures were 42.0%, 83.0%, 82.2%, 44.5%, and 48.0%, respectively.

DISCUSSION

We constructed a CNN for the automatic detection of protruding lesions in WCE small-bowel images. The trained CNN was shown to detect protruding lesions in independent test images with a sensitivity of 90.7% (AUC, 0.911). The detection rate of protruding lesions per patient was 98.6%. A variety of computer-aided diagnosis systems in WCE of the small bowel have been reported,[15-18] as have automated diagnosis systems using CNNs.[10,19-21] Our study was, to our best knowledge, the first to classify protruding lesions and to train and test a CNN using small-bowel images based on clinical diagnosis. Our CNN used more than 30,000 images and performed at a high level for the detection of protruding lesions. It is also interesting to have detected 2 potentially missing images from normal images, and it is surprising that these images were processed in as little as 0.0303 seconds per image.

To date, studies of AI-assisted diagnosis systems for images of the small bowel in WCE have included only a limited number of types of protruding legions, such as polyps.[10,21]

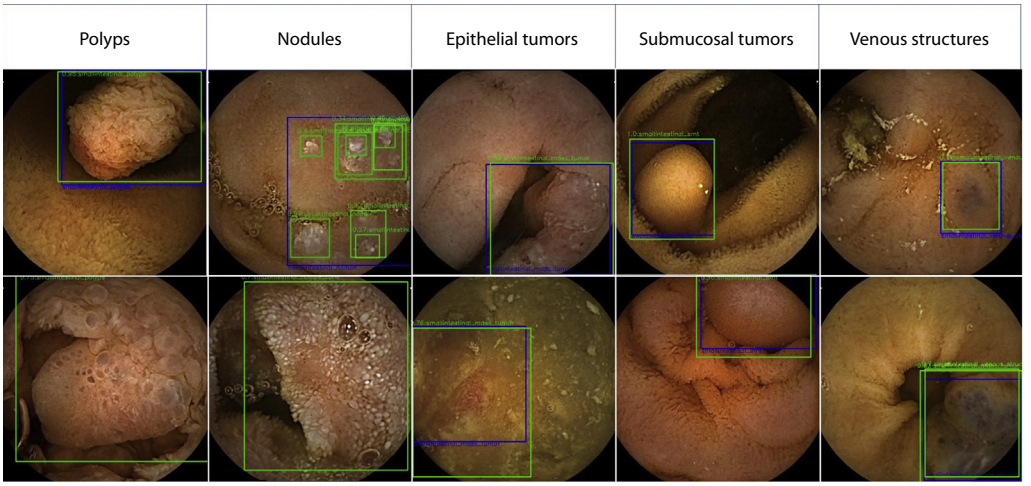


Figure 2. Representative images of protruding lesions correctly detected by the convolutional neural network (CNN) in the validation set (*blue box,* true lesion; *green box*, region identified as a protruding lesion by the CNN; number, the probability score determined by the CNN). The *blue* rectangular bounding boxes show the annotation of protruding regions as determined by experts. The *green* rectangular bounding boxes were applied by the CNN, with the name of lesion subcategory and its probability score.

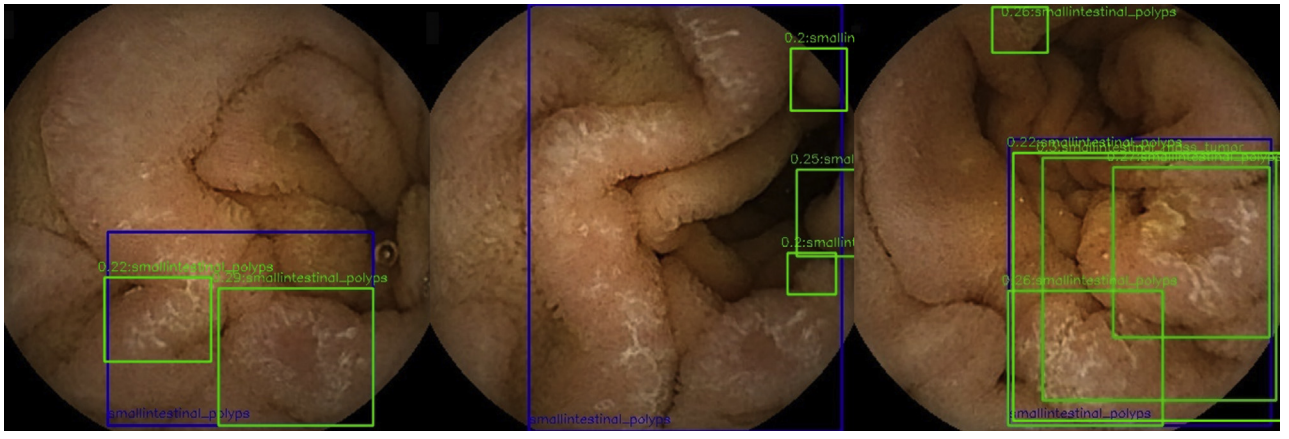


Figure 3. Protruding lesions that the convolutional neural network (CNN) could not detect. All polyps of 1 patient with false-negative results, with familial adenomatous polyposis (FAP), that the CNN could not detect. The *blue* rectangular bounding boxes show the annotation of protruding regions by experts, and the *green* rectangular bounding boxes are shaped by the CNN with the name of subcategories and its probability score. Every probability score attached to the *green* rectangular bounding boxes is <0.317; therefore, these images are not detected as images with protruding regions by the CNN.

Our study included not only polyps but a variety of other types of lesions, and the sensitivity of diagnosis of polyps was equivalent to that in a previous report (89%–95%).[21] Our study applied a category based on CEST, and it also revealed differences in sensitivity between categories, ie, polyps, nodules, epithelial tumors, SMTs, and venous structures. Compared with those in other parts of the GI tract, such as the colon, stomach, and esophagus, protruding lesions in the small intestine vary in morphology, depending on the disease, which makes automatic diagnosis difficult. Appropriate CNN training with classification of the protruding lesions appears to be the key for improving the CNN, as does a large number of collected images. In the present study, the CNN correctly classified more than 80% of epithelial tumors and nodules, whereas protruding legions of other categories were mistakenly identified as belonging to different categories. The CNN has a high sensitivity for detecting epithelial tumors, which suggests that the correct category classification may lead to improved sensitivity. Venous structures and SMTs have an insufficient number of training images compared with other categories, and a greater number of training images can be expected to improve performance. In addition, variations in lesion appearance within a category may be a significant hindrance. In particular, nodules vary in morphologic appearance. It may be difficult to regard images with a single nodule and images with a nodule gathering that occupies the whole image as being in the same category. It seems to be necessary to train the CNN on detailed subcategories according to the CEST classification.

TABLE 4. Causes of false-negative or false-positive images by the convolutional neural network (CNN)

Cause	n (%)
False-negative images (n = 697)	
Similar color to the surrounding normal mucosa	277 (39.7)
Partialness or laterality	146 (20.9)
Smallness*	96 (13.8)
Poorly demarcated†	
Darkness	85 (12.2)
Debris or foam	79 (11.3)
Poor focus	14 (2.0)
False-positive images (n = 2019)	
Normal mucosa	1844 (91.3)
Foam	132 (6.5)
Debris	37 (1.8)
Vascular dilatation	6 (0.3)

*The occupancy ratio of lesion in the whole image was small.
†Due to darkness, debris, foam, or poor focus.

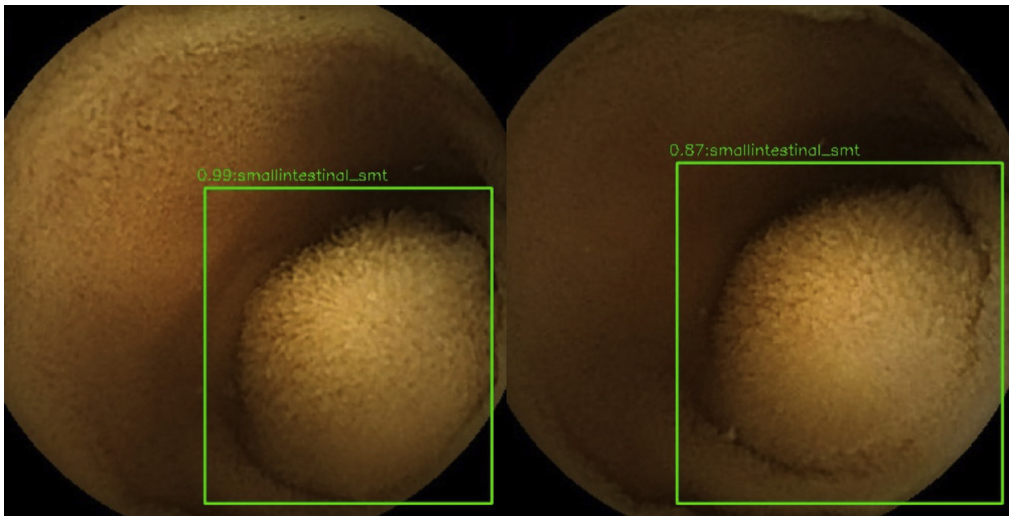


Figure 4. Images that were originally diagnosed as normal by experts but were re-evaluated as possible submucosal tumors (SMTs) by experts after evaluation by the convolutional neural network (CNN). The *green* rectangular bounding boxes were applied by the CNN, with the name of lesion subcategory and its probability score. For both images, the CNN applied *green* rectangular bounding boxes indicating SMTs with probability scores of 0.99 and 0.87, respectively.

Individual patient analyses constituted a novel feature of this study. Regrettably, the detection rate for protruding lesions per patient was not perfect. The polyps of the false-negative "patient" image had atypical morphology, and their cause was FAP (Fig. 3). This indicated that more cases of FAP are required for both training and testing to improve the CNN.

We reviewed all images that were incorrectly judged by the CNN. Two major reasons for false-negative images were the partialness and the similar colors of the protruding lesions to the surrounding healthy mucosa, and these might cause the relatively low sensitivity for polyps and SMT. On the other hand, the major reason of false-positive images was the normal mucosa that was mainly tagged as polyps or SMT by the CNN. These results indicate a dilemma that when the CNN is trained by more images of polyps and SMTs for improving its sensitivity for these images, the false-positive images of normal mucosa may increase. This point should be considered at further training and validation.

The study has several limitations. First, it was a retrospective study using only selected images. It is necessary to verify the accuracy of the present CNN on images in a clinical setting. Second, the number of patients in the test set was relatively small, although the patient-level analysis revealed high sensitivity of the CNN in detecting protruding lesions. The causes of protruding lesions used in this study are skewed, which reflects the frequency of

TABLE 5. Labeling of the protruding lesions by the experts and the CNN based on capsule endoscopy structured terminology (CEST) classification

Classification by the CNN	Experts' diagnoses (n = 7507)					Total
	Polyps	Nodules	Epithelial tumors	Submucosal tumors	Venous structures	
Polyps	1478 (42.0)	52 (2.7)	121 (8.3)	48 (14.2)	2 (0.8)	1701
Nodules	432 (12.3)	1604 (83.0)	39 (2.7)	2 (0.6)	0 (0)	2077
Epithelial tumors	1298 (36.9)	15 (0.8)	1202 (82.2)	53 (15.6)	86 (34.1)	2654
Submucosal tumors	31 (0.9)	0 (0)	39 (2.7)	151 (44.5)	29 (11.5)	250
Venous structures	0 (0)	0 (0)	0 (0)	7 (2.1)	121 (48.0)	128
No lesions	283 (8.0)	261 (13.5)	61 (4.2)	78 (23.0)	14 (5.6)	697
Total	3522	1932	1462	339	252	7507

In the case of an image not containing any regions with a probability score of $\geq$0.317, the CNN recognized that the image included no protruding lesions.
CNN, Convolutional neural network.

lesions seen clinically in Japan, and an international validation study would be needed. Third, only the Pillcam system was used in this study. Further research is required to verify the generalizability, using full videos of a large number of cases, based on various types of WCE systems. Finally, the present study focused on protruding lesions. The AI diagnostic system for mucosal breaks[9] and angioectasia[11] in our previous reports will be integrated and validated in the future.

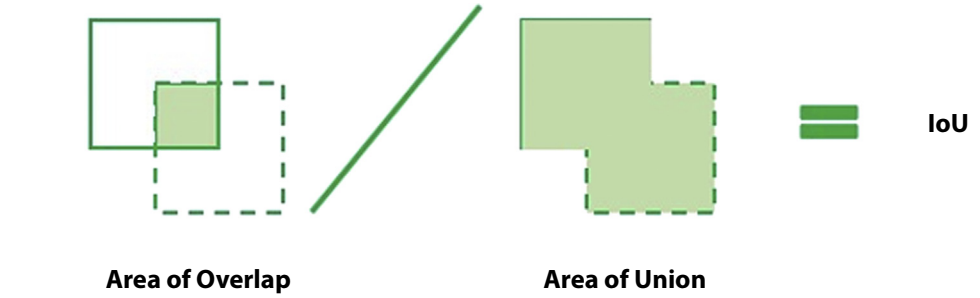
In conclusion, we have successfully trained a CNN to detect protruding lesions in small-bowel capsule endoscopy. Although the results of the current study have demonstrated sufficient capacity to detect lesions on a patient basis, further studies are necessary to achieve better diagnostic performance.

ACKNOWLEDGMENT

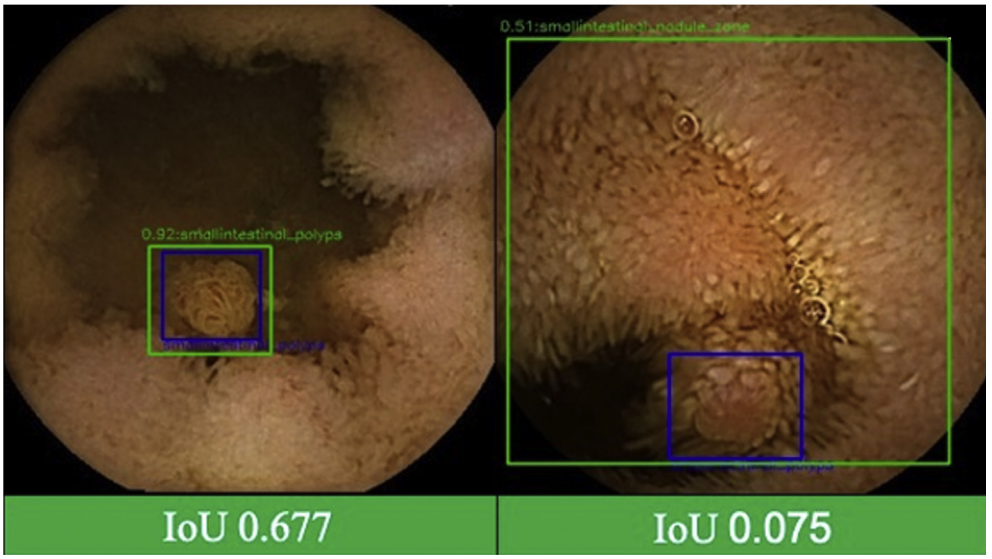
The authors thank the engineers at AI Medical Service Inc. (Tokyo, Japan) for their cooperation in developing and testing the CNN. They also thank Editage (www.editage.jp) for English language editing.

REFERENCES

1. Iddan G, Meron G, Glukhovsky A, et al. Wireless capsule endoscopy. Nature 2000;405 417-17.
2. Costamagna G, Shah S, Riccioni M, et al. A prospective trial comparing small bowel radiographs and video capsule endoscopy for suspected small bowel disease. Gastroenterology 2002;123:999-1005.
3. Omori T, Hara T, Sakasai S, et al. Does the PillCam SB3 capsule endoscopy system improve image reading efficiency irrespective of experience? A pilot study. Endosc Int Open 2018;6:E669-75.
4. McAlindon ME, Ching H-L, Yung D, et al. Capsule endoscopy of the small bowel. Ann Transl Med 2016;4:369.
5. Leenhardt R, Vasseur P, Li C, et al. A neural network algorithm for detection of GI angiectasia during small-bowel capsule endoscopy. Gastrointest Endosc 2019;89:189-94.
6. Korman LY, Delvaux M, Gay G, et al. Capsule endoscopy structured terminology (CEST): proposal of a standardized and structured terminology for reporting capsule endoscopy procedures. Endoscopy 2005;37:951-9.
7. Shah P, Kendall F, Khozin S, et al. Artificial intelligence and machine learning in clinical development: a translational perspective. NPJ Digit Med 2019;2:69.
8. Misawa M, Kudo SE, Mori Y, et al. Artificial intelligence-assisted polyp detection for colonoscopy: initial experience. Gastroenterology 2018;154:2027-9.
9. Aoki T, Yamada A, Aoyama K, et al. Automatic detection of erosions and ulcerations in wireless capsule endoscopy images based on a deep convolutional neural network. Gastrointest Endosc 2019;89:357-63 e2.
10. Ding Z, Shi H, Zhang H, et al. Gastroenterologist-level identification of small bowel diseases and normal variants by capsule endoscopy using a deep-learning model. Gastroenterology 2019;157:1044-54.
11. Tsuboi A, Oka S, Aoyama K, et al. Artificial intelligence using a convolutional neural network for automatic detection of small-bowel angioectasia in capsule endoscopy images. Dig Endosc. Epub 2019 Aug 7.
12. Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector. arXiv:1512.02325v5 [cs.CV] 2016. Available at: https://arxiv.org/pdf/1512.02325.pdf. Accessed October 29, 2019.
13. Yangqing J, Evan S, Jeff D, et al. Caffe: convolutional architecture for fast feature embedding. Available at: https://arxiv.org/pdf/1408.5093.pdf. Accessed October 29, 2019.
14. Youden WJ. Index for rating diagnostic tests. Cancer 1950;3:32-5.
15. Vieira PM, Freitas NR, Valente J, et al. Automatic detection of small bowel tumors in wireless capsule endoscopy images using ensemble learning. Med Phys 2019;47:52-63.
16. Iakovidis DK, Koulaouzidis A. Automatic lesion detection in capsule endoscopy based on color saliency: closer to an essential adjunct for reviewing software. Gastrointest Endosc 2014;80:877-83.
17. Al-Shebani Q, Premaratne P, McAndrew DJ, et al. A frame reduction system based on a color structural similarity (CSS) method and Bayer images analysis for capsule endoscopy. Artif Intell Med 2019;94:18-27.
18. Ghosh T, Fattah SA, Wahid KA. CHOBS: color histogram of block statistics for automatic bleeding detection in wireless capsule endoscopy video. IEEE J Transl Eng Health Med 2018;6:1800112.
19. Hwang Y, Park J, Lim YJ, et al. Application of artificial intelligence in capsule endoscopy: where are we now? Clin Endosc 2018;51:547-51.
20. Alaskar H, Hussain A, Al-Aseem N, et al. Application of convolutional neural networks for automated ulcer detection in wireless capsule endoscopy images. Sensors (Basel) 2019;19:E1265.
21. Yuan Y, Meng MQ. Deep learning for polyp recognition in wireless capsule endoscopy images. Med Phys 2017;44:1379-89.

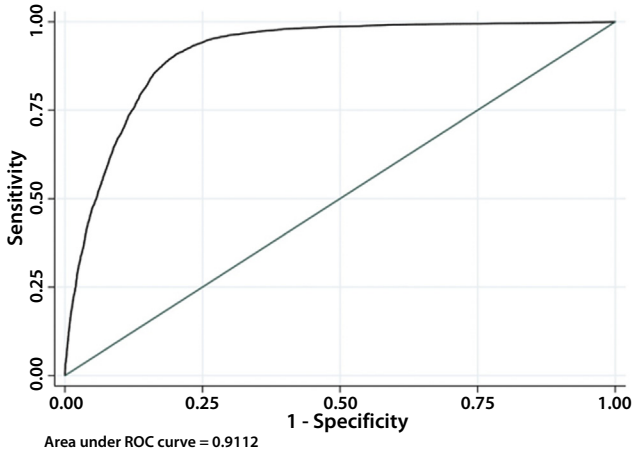


A



B C

Supplementary Figure 1. Visual description and examples of intersection over union (IoU). **A**, IoU is an evaluation method to measure the accuracy of an object detector; an IoU was calculated by dividing the area of overlap of 2 boxes by the area of union. The higher the IoU, the greater the overlap, and the better the result. We defined the CNN box with ≥0.05 IoU as overlapped. **B**, The image included a polyp. The polyp was attached with the *blue* square boxes by endoscopist as a "true" box, and the CNN enclosed the polyp by *green* square boxes. The 2 boxes showed the IoU of 0.677 (correct answer). **C**, The image included a nodule. The IoU of the 2 boxes were 0.075 (correct answer).



Supplementary Figure 2. The receiver operator characteristic (ROC) curve of the convolutional neural network (CNN) for detecting protruding lesions. *AUC,* area under the receiver operating characteristic curve; *CI,* Confidence interval.