Gastroenterologist-Level Identification of Small-Bowel Diseases and Normal Variants by Capsule Endoscopy Using a Deep-Learning Model

Zhen Ding,[1,*] **Huiying Shi**,[1,*] Hao Zhang,[2] Lingjun Meng,[1] Mengke Fan,[1] Chaoqun Han,[1] Kun Zhang,[1] Fanhua Ming,[2] Xiaoping Xie,[1] Hao Liu,[2] Jun Liu,[1] Rong Lin,[1] and Xiaohua Hou[1]

[1]*Department of Gastroenterology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China; and* [2]*Ankon Medical Technologies Co, Ltd, Shanghai, China*

BACKGROUND & AIMS: Capsule endoscopy has revolutionized investigation of the small bowel. However, this technique produces a video that is 8–10 hours long, so analysis is time consuming for gastroenterologists. Deep convolutional neural networks (CNNs) can recognize specific images among a large variety. We aimed to develop a CNN-based algorithm to assist in the evaluation of small bowel capsule endoscopy (SB-CE) images. **METHODS:** We collected 113,426,569 images from 6970 patients who had SB-CE at 77 medical centers from July 2016 through July 2018. A CNN-based auxiliary reading model was trained to differentiate abnormal from normal images using 158,235 SB-CE images from 1970 patients. Images were categorized as normal, inflammation, ulcer, polyps, lymphangiectasia, bleeding, vascular disease, protruding lesion, lymphatic follicular hyperplasia, diverticulum, parasite, and other. The model was further validated in 5000 patients (no patient was overlap with the 1970 patients in the training set); the same patients were evaluated by conventional analysis and CNN-based auxiliary analysis by 20 gastroenterologists. If there was agreement in image categorization between the conventional analysis and CNN model, no further evaluation was performed. If there was disagreement between the conventional analysis and CNN model, the gastroenterologists re-evaluated the image to confirm or reject the CNN categorization. **RESULTS:** In the SB-CE images from the validation set, 4206 abnormalities in 3280 patients were identified after final consensus evaluation. The CNN-based auxiliary model identified abnormalities with 99.88% sensitivity in the per-patient analysis (95% confidence interval [CI], 99.67–99.96) and 99.90% sensitivity in the per-lesion analysis (95% CI, 99.74–99.97). Conventional reading by the gastroenterologists identified abnormalities with 74.57% sensitivity (95% CI, 73.05–76.03) in the per-patient analysis and 76.89% in the per-lesion analysis (95% CI, 75.59–78.15). The mean reading time per patient was 96.6 ± 22.53 minutes by conventional reading and 5.9 ± 2.23 minutes by CNN-based auxiliary reading ($P < .001$). **CONCLUSIONS:** We validated the ability of a CNN-based algorithm to identify abnormalities in SB-CE images. The CNN-based auxiliary model identified abnormalities with higher levels of sensitivity and significantly shorter reading times than conventional analysis by gastroenterologists. This algorithm provides an important tool to help gastroenterologists analyze SB-CE images more efficiently and more accurately.

Keywords: Artificial Intelligence; Imaging; Intestine; Lesion.

The small bowel (SB) is difficult to examine by traditional endoscopic and radiologic techniques, and it has thus been seen as the black box of the gastrointestinal tract in the past few decades.[1,2] In recent years, the introduction of capsule endoscopy (CE) has revolutionized the diagnosis, monitoring, and management of SB diseases.[1] CE can be used to clearly observe abnormalities of the SB, such as erosions, ulcerations, angiodysplasias, petechiae, venectasias, lymphangiectasia, erythema, edema, changes of the villi, and external constrictions,[2] and it has been recommended by the current European guidelines for patients with obscure gastrointestinal bleeding (OGIB), suspected Crohn's disease with negative ileocolonoscopy result, suspected SB tumors, and inherited polyposis syndromes.[3,4] However, the major limitation of CE is the high time cost of processing the 8–10 hours of video data and reporting the results (time cost of 1–2 hours per case for a gastroenterologist).[5–7]

Recently, deep-learning algorithms, such as convolutional neural networks (CNNs), have been shown to exceed human performance in visual tasks and have been widely used to enable the extraction of highly representative features.[8,9] A deep-learning–based artificial intelligence (AI) model has been reported to have similar performance to dermatologists in the classification of skin cancers.[9] Moreover, a deep-learning model has been validated in the real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered standard colonoscopy videos.[10] However, a deep-learning model has not yet been established in the identification of SB-CE abnormalities.

*Authors share co-first authorship.

Abbreviations used in this paper: AI, artificial intelligence; CE, capsule endoscopy; CI, confidence interval; CNN, convolutional neural network; OGIB, obscure gastrointestinal bleeding; SB, small bowel.

0016-5085/$36.00
https://doi.org/10.1053/j.gastro.2019.06.025

WHAT YOU NEED TO KNOW

BACKGROUND AND CONTEXT

Deep convolutional neural networks (CNNs) can recognize specific images among a large variety.

NEW FINDINGS

We validated the ability of a CNN-based algorithm to identify abnormalities in SB-CE images. The CNN-based auxiliary model identified abnormalities with higher levels of sensitivity and significantly shorter reading times than conventional analysis by gastroenterologists.

LIMITATIONS

This system requires evaluation in a large, prospective study.

IMPACT

This algorithm provides an important tool to help gastroenterologists analyze SB-CE images more efficiently and more accurately.

In this study, we used CNN to train a deep-learning–based AI model to differentiate abnormal images from normal images in SB-CE examinations and validated the algorithm model using a large multicenter data set.

Methods

Study Design

This study was performed in 77 participating medical examination centers with the ESView platform (developed by Ankon Technologies Co, Ltd, Shanghai, China). Data for 6970 patients (113,426,569 images) having SB examination with CE in these centers were collected between July 2016 and July 2018: 1970 cases were used to establish the CNN-based auxiliary reading model in the training phase, and 5000 cases were used to validate the CNN-based auxiliary reading model in the validation phase. There was no patient overlap between the 2 groups. Written informed consent was obtained from all patients. Patient data were anonymized, and any personally identifying information was omitted. This study was approved by the Ethics Committee of Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China.

Because the purpose of our CNN-based system design is to ensure the highest sensitivity—that is, to screen out as many lesions as possible—the abnormal images in this study were defined as 2 different categories: clinically significant abnormal lesions (abnormal lesions, eg, inflammation, ulcer, polyps, protruding lesion, vascular disease, bleeding, parasite, and diverticulum) and normal variants (lymphangectasia, lymphoid hyperplasia, etc).

Capsule Endoscopy System and Procedure

Patients were examined with CE (Ankon Medical Technologies Co, Ltd). The Ankon CE system consists of 3 components: an endoscopic capsule, a data recorder, and a computer workstation with software for real-time viewing and controlling. The capsule is 27 mm in length and 11.8 mm in diameter; it weighs 4.8 g. It has a single camera, and the field of view is larger than 140° ± 10%. The capsule passes passively through the SB, and the dynamic frame rate is 0–2 frames/s. Images are captured and recorded at an average rate of 0.8 frames/s with a resolution of 480 × 480. Each video consists of a continuous single image or frame. Each image or frame in a video was tagged with a specific number, in the order that the image was taken, and saved to a folder. The battery life of the capsule is ≥8 hours.

All patients received a bowel preparation with polyethylene glycol electrolyte solution, fasted overnight, and arrived at the centers for CE examination in the morning. At 30 minutes before CE examination, patients ingested 30 mL simethicone suspension (Espumisan, 40 mg/mL; Berlin-Chemie, Germany) dissolved in 50 mL water; a small amount of water was then used to swallow the capsule, as previously reported.[11] The SB cleansing score used in this study was based on previously published studies,[12,13] and the SB image quality results of the patients enrolled in this study were evaluated (Supplementary Table 1). We used ESNavi (a cloud platform developed by Ankon Technologies Co, Ltd) as a remote reading system, enabling remote examination, reading, and data storage and sharing.

The Training Phase of a Convolutional Neural Network–Based Auxiliary Reading Model

We developed a deep-learning–based AI model to identify abnormal images from normal images in the SB-CE examination. This included 2 phases: a training phase and a validation phase (Figure 1).

In the training phase, we trained a CNN-based algorithm—a type of artificial neural network used in deep learning and a method for the analysis of visual imagery—to recognize abnormal SB images. First, the model was trained by data from ImageNet (a large visual database designed for use in visual object recognition software) to obtain a preprocessing model (Figure 1). Second, labeled SB-CE image data (indicating whether or not the image had a lesion) was used for training by the residual neural network (ResNet) model[14] based on the preprocessing model (Figure 1). In detail, we selected normal images and abnormal images (including 10 categories: inflammation, ulcer, polyps, lymphangiectasia, bleeding, vascular disease, protruding lesion, lymphatic follicular hyperplasia, diverticulum, and parasite) to train the model. To ensure that at least 1000 representative images for each abnormal category were analyzed, the 1970 patients were chosen as the training group (a total of 158,235 images were totally selected by 2 experienced gastroenterologists from Wuhan Union Hospital). As a result, the CNN model parameters were updated to obtain a trained CNN model for the assessment of SB-CE images.

The CNN-based system would calculate, for each image, the probability that it was an abnormal image according to the probability threshold value set by the model (Figure 2*A*). The structure of residual learning of the CNN-based system is shown in the Figure 2*B*, which called "shortcut connections."[14] The probability threshold value (sensitivity, 95.4%; specificity, 96.99%) of the CNN-based system used in this study was set based on the receiver operating characteristic curve using the Youden index method (Figure 2*C*). The confusion matrix of the CNN-based AI model in the training phase is shown in Supplementary Table 2.

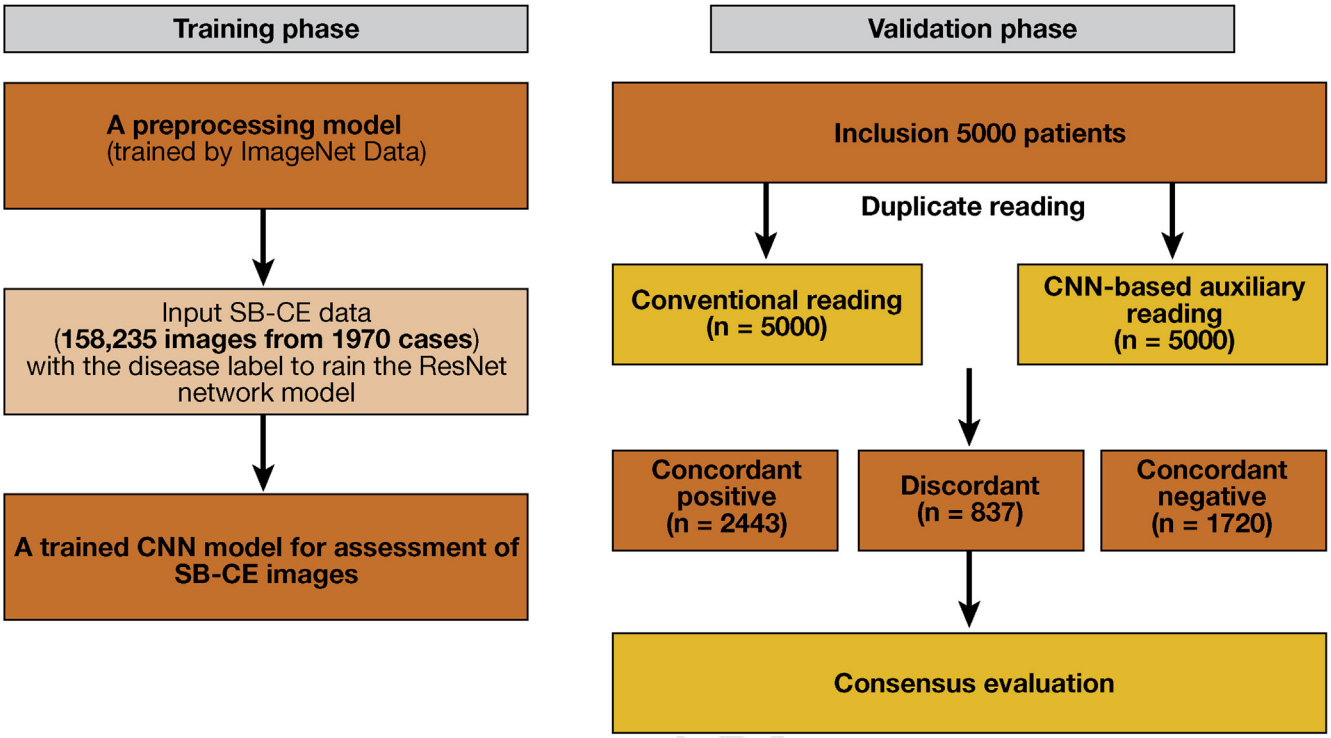


Figure 1. Study flow charts of the training and validation phases.

The Validation Phase of a Convolutional Neural Network–Based Auxiliary Reading Model

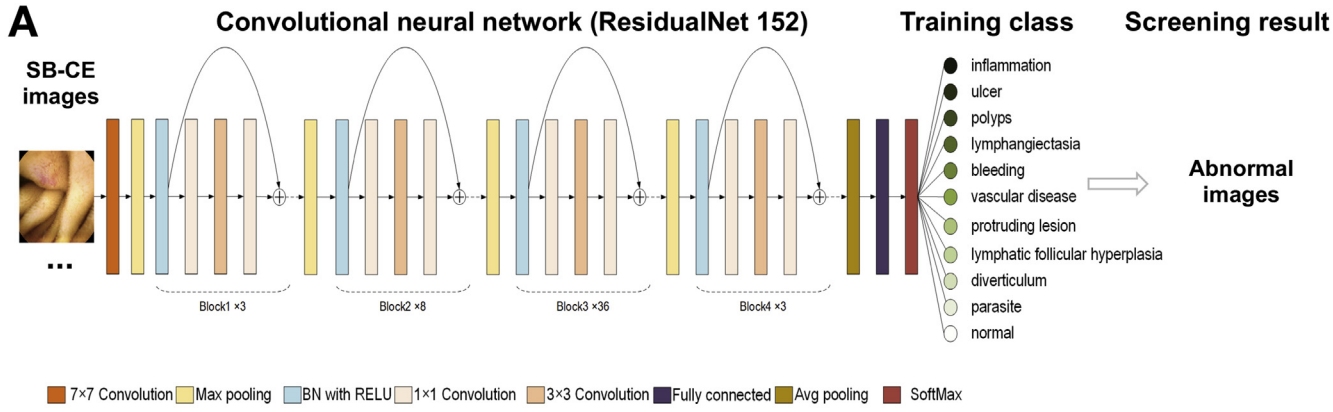
In the validation phase, all 5000 recordings (113,268,334 images) were read by both conventional reading and CNN-based auxiliary reading by 20 gastroenterologists who routinely perform SB-CE examination and evaluation in clinical practice.

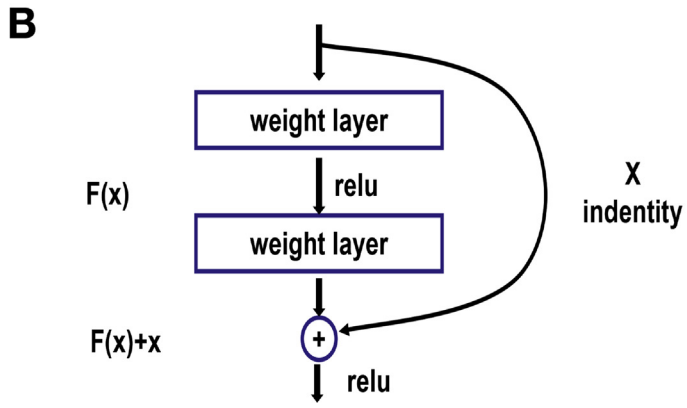
For conventional reading, all 5000 videos from the 5000 patients were read by the team of 20 gastroenterologists. These videos were randomly and equally distributed to the 20 gastroenterologists, and each gastroenterologist received 250 videos. Each gastroenterologist read all original images in each video. For CNN-based auxiliary reading, all original SB-CE images from the 5000 patients were first input into the CNN-based auxiliary reading model, and suspected abnormal images auto-filtered by the model were further reviewed manually by gastroenterologists. The suspected abnormal images were selected by the CNN-based auxiliary reading model as described in the training phase. All gastroenterologists independently evaluated all 250 cases and made a diagnosis. Diagnoses and the total time for each SB-CE image evaluation were recorded.

When a diagnostic agreement was reached between conventional and CNN-based auxiliary reading, no further evaluation was carried out. In the case of a discordant final diagnosis and/or different lesions observed, the 20 gastroenterologists sat together, and the images of the patient were reevaluated to confirm or reject the discordance. Only the final consensus diagnoses were considered as the reference standard of diagnosis. For lesions identified by conventional reading that were not detected by CNN-based auxiliary reading in a patient, we checked through auto-filtered suspected abnormal images to determine whether the CNN-based auxiliary reading model had failed to detect the lesion. For lesions identified by CNN-based auxiliary reading that were not detected by conventional reading in a patient, we reexamined the filtered suspected abnormal images identified by CNN-based auxiliary reading and original videos of that patient. Because each suspected abnormal image identified by the CNN-based auxiliary system in a given patient was specifically tagged, it could be easily traced to the location in the original video for this patient.

Statistical Analysis

Diagnoses made by gastroenterologists were sorted and counted according to the following categories: normal, inflammation, ulcer, polyps, lymphangiectasia, bleeding, vascular disease, protruding lesion, lymphatic follicular hyperplasia, diverticulum, parasite, and other, both per patient and per lesion. All relevant data were entered into a customized database and then analyzed with SPSS software, version 21.0 (IBM, Armonk, NY). A chi-squared test was performed to analyze the difference in detection rates between conventional and CNN-based auxiliary reading for each specific SB disease classification. An independent-sample t test was used to compare the total time spent to evaluate the SB-CE images between conventional reading and CNN-based auxiliary reading. The McNemar test was applied to compare the diagnostic performance between conventional reading and CNN-based auxiliary reading using the consensus evaluation review as the reference standard; all McNemar tests were analyzed with SAS software, version 9.4 (SAS Institute Inc., Cary, NC). The Cochran-Mantel-Haenszel test was used to analyze the interobserver agreement (κ) among the 20 gastroenterologists in per-patient and per-lesion analyses, using SAS software, version 9.4. Sensitivity, specificity, positive predictive values, and negative predictive values were described as percentage and 95% confidence interval (CI). A value of $P < .05$ was considered statistically significant.





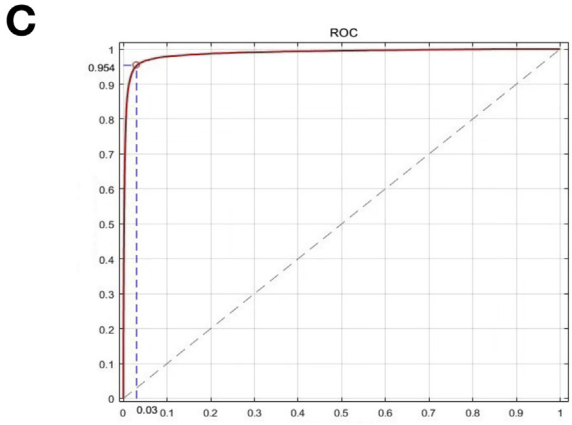


Figure 2. CNN-based algorithm reading model. (*A*) Process for detection of abnormal SB-CE images. Data flow is from left to right: an image of SB-CE is subsequently processed with the CNN-based model (ResidualNet), and images of suspicious anomalies are finally filtered out and output. (*B*) The structure of residual learning of CNNs. (*C*) Receiver operating characteristic (ROC) curve of the CNN-based algorithm in recognition of SB-CE lesion images.

Results

Detection of Small-Bowel Disease by Capsule Endoscopy Using the Convolutional Neural Network–Based Auxiliary Reading Model

A total of 5000 patients having SB-CE examination were enrolled in this study. The CNN-based auxiliary reading model can detect most of the abnormal SB images, including inflammation, ulcer, polyps, lymphangiectasia, bleeding, vascular disease, protruding lesion, lymphatic follicular hyperplasia, diverticulum, parasite, and other. Representative abnormal SB images, including endoscopic and CNN-processed images, are shown in Figure 3.

High Sensitivity of the Convolutional Neural Network–Based Auxiliary Reading Model in the Detection of Small-Bowel Abnormalities

Diagnoses of SB abnormalities with conventional reading and CNN-based auxiliary reading, per patient and per lesion, were analyzed according to consensus evaluation (Supplementary Tables 3 and 4). Overall, in the per-patient analysis, gastroenterologists achieved a sensitivity of 74.57% (95% CI, 73.05–76.03) in conventional reading and a sensitivity of 99.88% (95% CI, 99.67-99.96) in CNN-based auxiliary reading ($P < .0001$) (Table 1). Conventional reading missed diagnoses in 834 patients, and CNN-based auxiliary reading missed diagnoses in 4 patients. In the per-lesion analysis, gastroenterologists achieved a sensitivity of 76.89% (95% CI, 75.58–78.15) in conventional reading and a sensitivity of 99.90% (95% CI, 99.74–99.97) in CNN-based auxiliary reading ($P < .0001$) (Table 1). Conventional reading missed 972 abnormal diagnoses, and CNN-based auxiliary reading missed 4 abnormal diagnoses: 2 vascular diseases, 1 ulcer, and 1 protruding lesion.

When the secondary per-lesion analysis of clinically significant lesions and normal variants was performed, results showed that the sensitivity for clinically significant lesions was 88.02% (95% CI, 86.75–89.19) by conventional reading compared with 99.86% (95% CI, 99.61–99.95) by CNN-based auxiliary reading and that the sensitivity for normal variants was 54.98% (95% CI, 52.34–57.58) by conventional reading compared with 100% (95% CI, 99.66–100) by CNN-based auxiliary reading ($P < .0001$ for both) (Table 2).

Furthermore, the sensitivity and specificity of the conventional reading and the CNN-based auxiliary reading classified by abnormal SB image types in both per-patient and per-lesion analyses were calculated (Table 2).

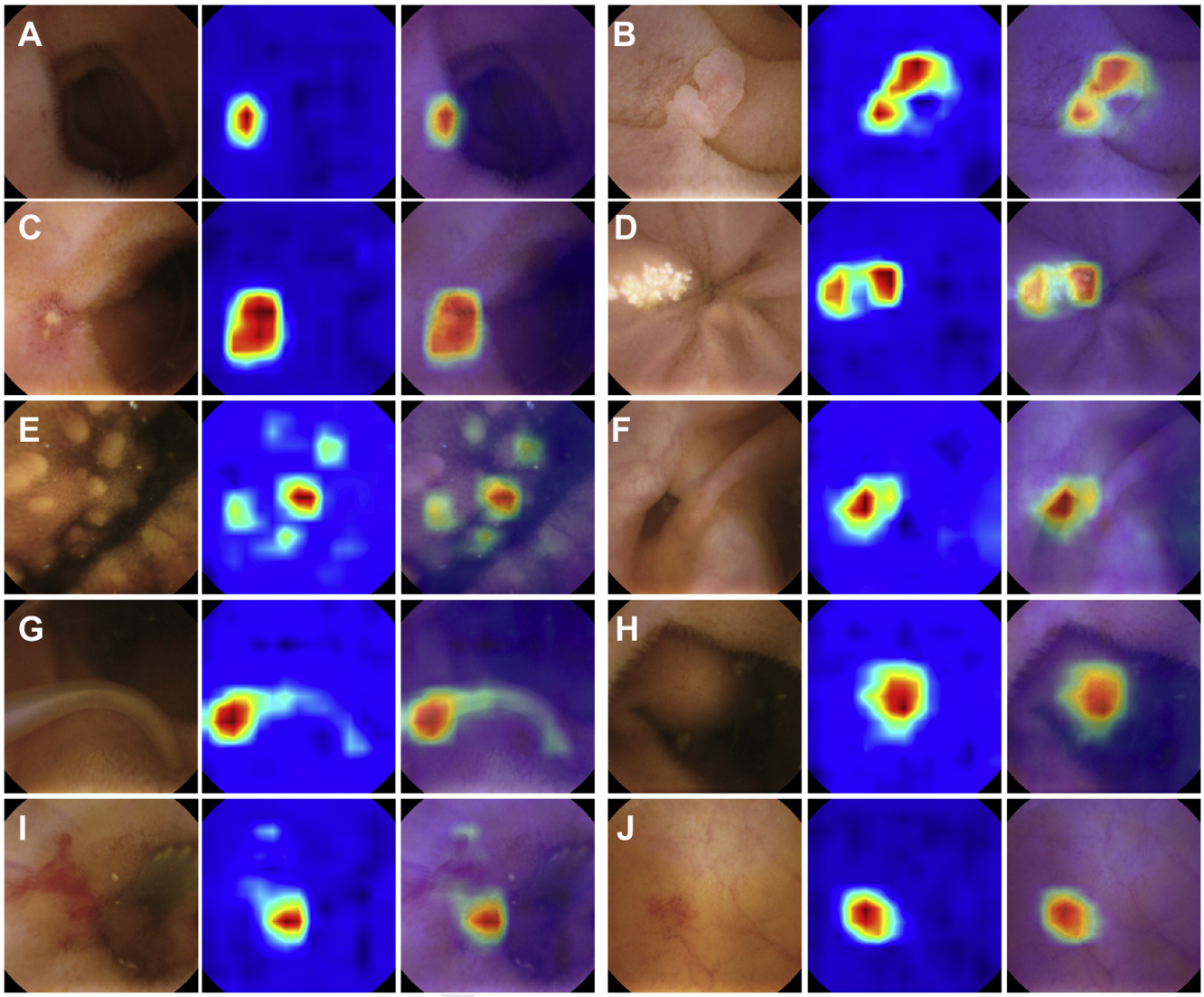


Figure 3. Representative capsule endoscopy images of SB abnormalities, including original endoscopic images and CNN-processed images: (*A*) inflammation, (*B*) polyps, (*C*) ulcer, (*D*) lymphangiectasia, (*E*) lymphatic follicular hyperplasia, (*F*) diverticulum, (*G*) parasite, (*H*) protruding lesion, (*I*) bleeding, and (*J*) vascular disease.

Table 1. Diagnostic Performance of Conventional Reading Group and CNN-Based Auxiliary Reading Group in Per-Patient and Per-Lesion Analyses

	Conventional reading group, % (95% CI)	CNN-based auxiliary reading group, % (95% CI)	*P* value
Per-patient analysis			
Sensitivity	74.57 (73.05–76.03)	99.88 (99.67–99.96)	<.0001
Specificity	100 (99.72–100)	100 (99.72–100)	1
PPV	100 (99.80–100)	100 (99.85–100)	1
NPV	67.35 (65.48–69.16)	99.77 (99.36–99.92)	<.0001
Per-lesion analysis			
Sensitivity	76.89 (75.58–78.15)	99.90 (99.74–99.97)	<.0001
Specificity	100 (99.72–100)	100 (99.72–100)	1
PPV	100 (99.85–100)	100 (99.89–100)	1
NPV	63.89 (62.04–65.70)	99.77 (99.36–99.93)	<.0001

NPV, negative predictive value; PPV, positive predictive value.

Table 2. Sensitivity and Specificity of Conventional and CNN-Based Auxiliary Readings According to SB Abnormal Lesions or Normal Variants in Per-Patient and Per-Lesion Analyses Q23

	Conventional reading, % (95% CI)	CNN-based auxiliary reading, % (95% CI)	*P* value
Per-patient analysis			
SB abnormal lesions			
Inflammation			
Sensitivity	93.87 (92.58–94.94)	100 (99.72–100)	<.0001
Specificity	100 (99.89–100)	100 (99.89–100)	1
Ulcer			
Sensitivity	98.12 (96.00–99.18)	99.73 (98.28–99.99)	.0339
Specificity	100 (99.91–100)	100 (99.91–100)	1
Polyps			
Sensitivity	78.16 (72.56–82.92)	100 (98.19–100)	<.0001
Specificity	100 (99.92–100)	100 (99.92–100)	1
Protruding lesion			
Sensitivity	56.14 (49.43–62.64)	99.56 (97.20–99.98)	<.0001
Specificity	100 (99.92–100)	100 (99.92–100)	1
Vascular disease			
Sensitivity	68.11 (60.80–74.65)	98.92 (95.74–99.81)	<.0001
Specificity	100 (99.92–100)	100 (99.92–100)	1
Bleeding			
Sensitivity	79.49 (63.06–90.13)	100 (88.83–100)	.0047
Specificity	100 (99.92–100)	100 (99.92–100)	1
Parasite			
Sensitivity	100 (69.87–100)	100 (69.87–100)	1
Specificity	100 (99.92–100)	100 (99.92–100)	1
Diverticulum			
Sensitivity	100 (69.87–100)	100 (69.87–100)	1
Specificity	100 (99.92–100)	100 (99.92–100)	1
Normal variants			
Lymphangiectasia			
Sensitivity	51.35 (47.94–54.75)	100 (99.44–100)	<.0001
Specificity	100 (99.91–100)	100 (99.91–100)	1
Lymphatic follicular hyperplasia			
Sensitivity	46.69 (40.50–52.99)	100 (98.16–100)	<.0001
Specificity	100 (99.92–100)	100 (99.92–100)	1
Other			
Sensitivity	71.99 (66.55–76.87)	100 (98.46–100)	<.0001
Specificity	100 (99.91–100)	100 (99.91–100)	1
Per-lesion analysis			
SB abnormal lesions			
Sensitivity	88.02 (86.75–89.19)	99.86 (99.61–99.95)	<.0001
Specificity	100 (99.85–100)	100 (99.85–100)	1
Inflammation			
Sensitivity	93.87 (92.58–94.94)	100 (99.72–100)	<.0001
Specificity	100 (99.86–100)	100 (99.86–100)	1
Ulcer			
Sensitivity	98.12 (96.00–99.18)	99.73 (98.28–99.99)	.0339
Specificity	100 (99.90–100)	100 (99.90–100)	1
Polyps			
Sensitivity	78.16 (72.56–82.92)	100 (98.19–100)	<.0001
Specificity	100 (99.90–100)	100 (99.90–100)	1
Protruding lesion			
Sensitivity	56.14 (49.43–62.64)	99.56 (97.20–99.98)	<0.0001
Specificity	100 (99.90–100)	100 (99.90–100)	1
Vascular disease			
Sensitivity	68.11 (60.80–74.65)	98.92 (95.74–99.81)	<.0001
Specificity	100 (99.90–100)	100 (99.90–100)	1
Bleeding			
Sensitivity	79.49 (63.06–90.13)	100 (88.83–100)	.0047
Specificity	100 (99.90–100)	100 (99.90–100)	1
Parasite			
Sensitivity	100 (69.87–100)	100 (69.87–100)	1
Specificity	100 (99.90–100)	100 (99.90–100)	1

Table 2. Continued

	Conventional reading, % (95% CI)	CNN-based auxiliary reading, % (95% CI)	*P* value
Diverticulum			
Sensitivity	100 (69.87–100)	100 (69.87–100)	1
Specificity	100 (99.90–100)	100 (99.90–100)	1
Normal variants			
Sensitivity	54.98 (52.34–57.58)	100 (99.66–100)	<.0001
Specificity	100 (99.89–100)	100 (99.89–100)	1
Lymphangiectasia			
Sensitivity	51.35 (47.94–54.75)	100 (99.44–100)	<.0001
Specificity	100 (99.89–100)	100 (99.89–100)	1
Lymphatic follicular hyperplasia			
Sensitivity	46.69 (40.50–52.99)	100 (98.16–100)	<.0001
Specificity	100 (99.90–100)	100 (99.90–100)	1
Other			
Sensitivity	71.99 (66.55–76.87)	100 (98.46–100)	<.0001
Specificity	100 (99.90–100)	100 (99.90–100)	1

Compared with the conventional reading group, the CNN-based auxiliary reading group showed a significant superiority in sensitivity for inflammation ($P < .0001$), ulcer ($P = .0339$), polyps ($P < .0001$), protruding lesion ($P < .0001$), vascular disease ($P < .0001$), bleeding ($P = .0047$), lymphangiectasia ($P < .0001$), and lymphatic follicular hyperplasia ($P < .0001$) detections and no difference in the sensitivity for parasite and diverticulum detections.

High Detection Rate of the Convolutional Neural Network–Based Auxiliary Reading Model in the Detection of Small-Bowel Abnormalities

Among 5000 patients, 3280 (a total of 4206 SB abnormal diagnoses) had an SB abnormal diagnosis after consensus evaluation review (Supplementary Tables 3 and 4). Furthermore, the detection rate of the conventional reading and the CNN-based auxiliary reading classified by abnormal SB image types both per patient and per lesion were analyzed (Supplementary Tables 5 and 6). In the per-lesion analysis, gastroenterologists detected 54.57% (n = 3234) of the total SB-CE diagnoses in conventional reading group, whereas the number was 70.91% (n = 4202) in the CNN-based auxiliary reading group (Supplementary Table 5) ($P < .0001$). Compared with conventional reading, CNN-based auxiliary reading had a higher detection rate in some types of SB abnormalities: increased detection rates were 7% in lymphangiectasia ($P < .0001$), 2.32% in lymphatic follicular hyperplasia ($P < .0001$), 1.74% in inflammation ($P = .034$), 1.67% in protruding lesion ($P < .0001$), 0.96% in polyps ($P = .007$), and 0.96% in vascular disease ($P = .001$). However, there was no significant difference in bleeding, ulcer, parasite, and diverticulum between these 2 groups.

A secondary per-lesion analysis of clinically significant lesions and normal variants was performed, and the results showed that the total detection rate increased 16.34% by CNN-based auxiliary reading compared with conventional reading (70.91% vs 54.57%), 5.57% for SB abnormal lesions, and 10.77% for normal variants (Supplementary Table 5 and Figure 4).

High Time Efficiency of the Convolutional Neural Network–Based Auxiliary Reading System in the Detection of Small-Bowel Abnormalities

With conventional reading, an average of 22,654 images were read for each patient, and an average of 578 images were identified as abnormal and filtered out for each patient with CNN-based auxiliary reading (Figure 5*A*). Moreover, the mean reading time was 96.6 ± 22.53 minutes with conventional reading, whereas the mean reading time with CNN-based auxiliary reading was 5.9 ± 2.23 minutes (Figure 5*B*) ($P < .001$).

Discussion

In this study, we explored a CNN-based algorithm model to assist with evaluation of SB-CE images and successfully validated the model with a large multicenter data set. Several highlights should be emphasized in our study. First, to our knowledge, this is the largest study assessing a CNN-based auxiliary reading model in SB-CE recordings with multicenter data. Second, to our knowledge, this is the first report of a CNN-based deep-learning model developed to assist with the diagnosis of multiple subtypes of abnormal SB lesions (including the normal variants) that can provide a new effective reading pattern for SB-CE. Third, we built a CNN-based algorithm with a high sensitivity (improving sensitivity from 76.89% to 99.90%), a high screening rate (reducing the number of images from 22,654 to 578), a high lesion detection rate (improving detection from 54.57% to 70.91%), and high time efficiency (reducing the average time required for reading from 96.6 minutes to 5.9 minutes) for detection of SB abnormalities using CE.

CE is increasingly performed in the clinical examination of SB diseases. The main limitation of CE is the long time required to review images. Deep-learning models have achieved excellent performance in image recognition of skin cancers[9] and real-time differentiation and detection of adenomatous and colorectal polyps[10] and gastrointestinal angiectasia.[5] Therefore, we aimed to develop a CNN-based

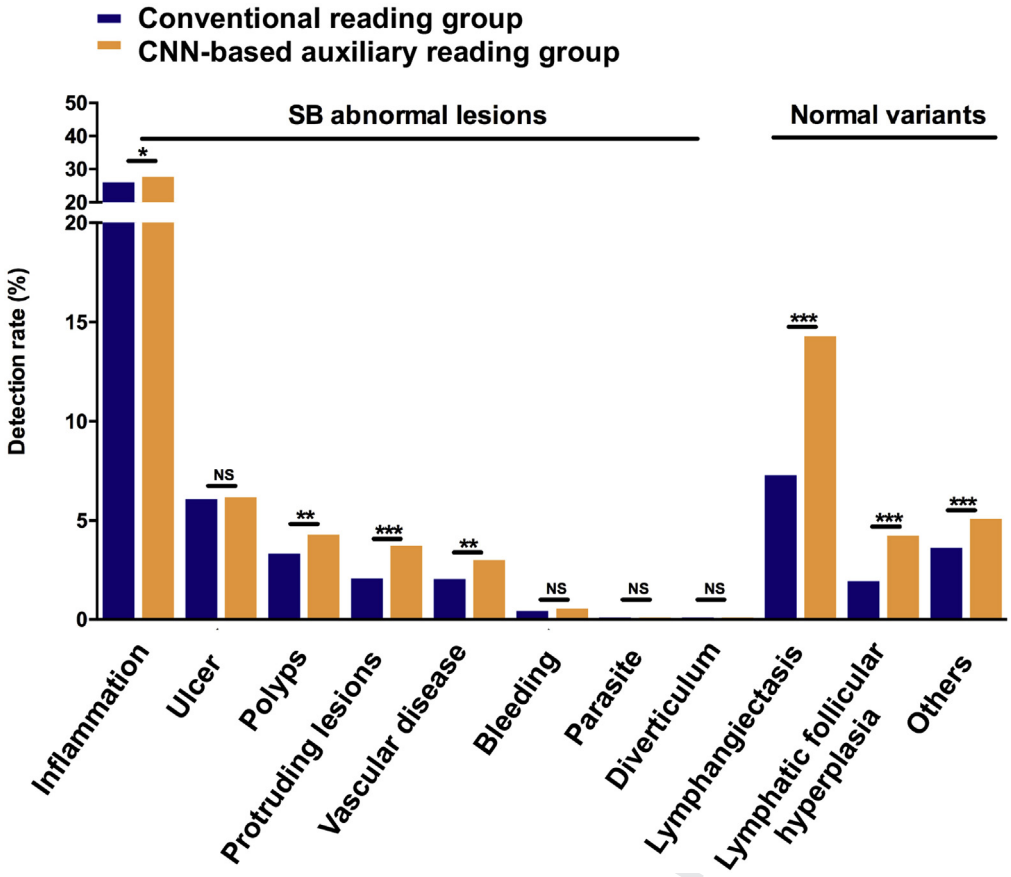


Figure 4. The detection rate for SB abnormal lesions or normal variants in the conventional reading group (*blue*) and CNN-based auxiliary reading group (*yellow*). *$P < .05$, **$P < .01$, ***$P < .0001$. NS, no significance.

auxiliary reading model for assessment of SB-CE images that would reduce the reading time.

In the past decade, several computer-based medical systems have been made to analyze CE images using algorithms[15]; however, to our knowledge, they have yet to yield a sufficient diagnostic performance to be considered for widespread clinical use.[5] A systematic review and meta-analysis verified the validity of suspected blood indicator software (Medtronic, Minneapolis, MN) in CE, but the overall performance for bleeding or potentially bleeding lesions was not satisfactory (sensitivity of 55.3%, specificity of 57.8%).[16] The Smart QuickView software could enable the user to view only the images that contain possible interesting areas, as detected by proprietary algorithms in the RAPID application, but results were discordant with conventional human reading in 28.3% of cases.[17] The result of multicenter prospective evaluation of the ExpressView (Capsovision, Saratoga, CA) reading mode for SB-CE studies showed that the mean reading time of capsule films was reduced and that a sensitivity of 78.6% for the detection of relevant lesions was achieved.[18]

Zhou et al[19] reported that a trained CNN-based model (GoogLeNet) was able to distinguish the frames from capsule endoscopy clips of patient with celiac disease patient vs control individuals, with a 100% sensitivity and specificity.[19] Leenhardt et al[5] developed an algorithm that associated a segmentation approach with a CNN-based approach for deep feature extraction, which yielded a sensitivity of 100.0% and a specificity of 96% for gastrointestinal angiectasia detection. However, this is not enough to solve the real clinical problem posed by CE: the detection of tens or even hundreds of abnormalities is necessary.[15] We believe that the main purpose of the work by Leenhardt et al and ours is quite different. Leenhardt et al's study developed a CNN-assisted diagnosis tool for the detection of 1 specific disease, gastrointestinal angiectasia, whereas the main purpose of our study was not limited to identifying only 1 specific disease; it aimed to develop an algorithm capable of differentiating multiple types of abnormal SB-CE images from normal images. This CNN-based algorithm could provide a novel timesaving tool for gastroenterologists to read SB-CE videos more effectively and accurately.

In our study, we showed that the trained CNN-based auxiliary reading model was able to detect and correctly identify most of the abnormal SB images, including inflammation, ulcer, polyps, lymphangiectasia, bleeding, vascular disease, protruding lesion, lymphatic follicular hyperplasia, diverticulum, and parasite. Of note, we showed that CNN-based algorithm reduced the reading time of SB-CEs by 93.9%, to 5.9 minutes. In a previous study, the ExpressView algorithm reduced the reading time of SB-CEs by 50%, from 39.7 minutes to 19.7 minutes,[18] and the mean reading time in QuickView mode was 11.6 minutes.[17] In terms of the overall diagnostic performance of SB abnormalities, a

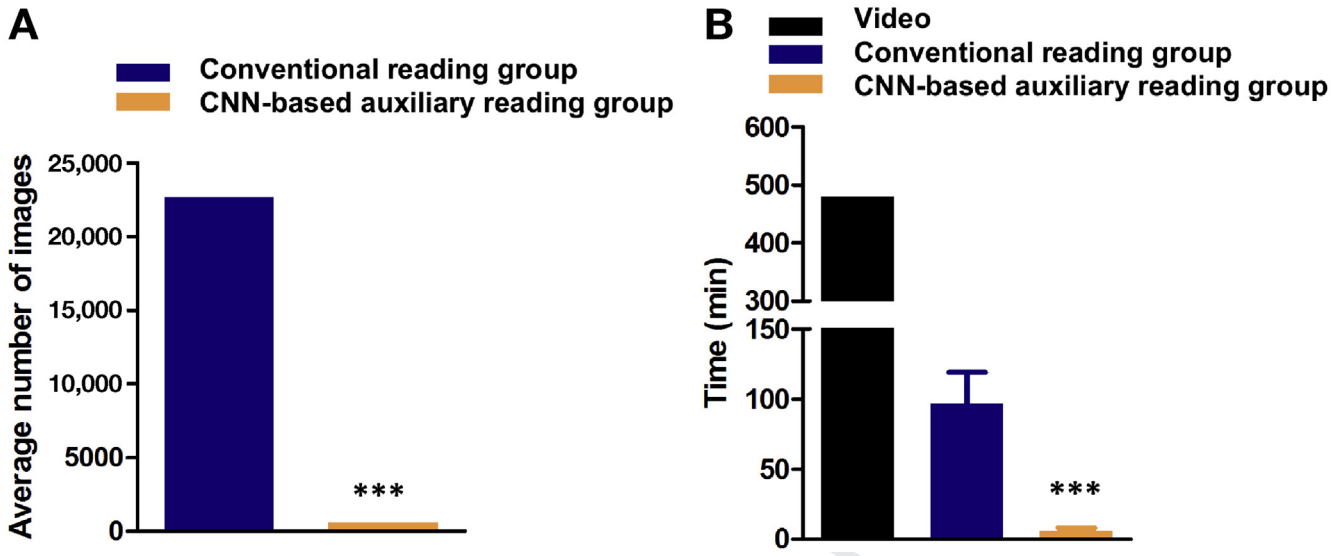


Figure 5. The average number and time for SB-CE reading in conventional reading and CNN-based auxiliary reading. (*A*) The average number of images read by conventional reading (*blue*) and by CNN-based auxiliary reading (*yellow*). (*B*) The total time of SB-CE examination video (*black*) and the mean reading time for the SB-CE images by the conventional reading group (*blue*) and by CNN-based auxiliary reading (*yellow*). ***$P < .0001$.

significant increase in sensitivity and lesion detection rate were found in the CNN-based auxiliary reading group compared with the conventional reading group. However, when a secondary analysis of clinically significant lesions and normal variants was performed, the results showed that the total detection rate increased 16.33% by the CNN-based auxiliary reading algorithm compared with the conventional group (70.90% vs 54.57%). For SB abnormal lesions, the detection rate increased 5.57%. For normal variants, the detection rate increased 10.77%.

The algorithm used in this study had an error rate of only 3% (ie, 3% of suspected abnormal images identified by CNN-based auxiliary were actually normal images). The reason for setting up this error recognition rate was that we were trying to achieve higher sensitivity for the CNN-based auxiliary system. In other words, we were trying to screen out as many suspected abnormal images as possible. Furthermore, we made the interobserver agreement (κ) among 20 gastroenterologists in their capsule interpretation in a per-patient and per-lesion analysis, and the results showed that the intraobserver agreement among the 20 gastroenterologists was high (κ scores of 0.783–0.811 in conventional reading mode and 0.937–0.944 in CNN-based auxiliary reading mode) (Supplementary Table 7).

As we all know, it has been well reported that an overall estimated colon polyp miss rate was more than 20%, with an increasing miss rate with decreased size of the polyp and adenoma during the colonoscopy.[20,21] A recent study from Wang et al[22] reported that an AI system significantly increased hyperplastic polyp detection rates from 35.07% to 64.93%, which is an increase of almost 30 percentage points, especially for polyps with a size of <5 mm.[22] For SB-CE reading, maintaining a high level of concentration while viewing 50,000–100,000 images of the same organ is difficult[23] and might entail an inherent risk of lesions being missed by physicians during the reading process.[5,23] There is evidence that some missed lesions present on the visual field during the reading but are not recognized by the endoscopist.

The most common CE diagnoses were bleeding/vascular diseases, inflammation, ulcer, and polyps.[24] The mucosal inflammation was referred to as the inflammatory changes of intestine under endoscopy, defined in previous publications, including erosion, Crohn's disease, NSAID enteritis, intestinal tuberculosis, radiation enteritis, eosinophilic enteritis, etc.[24,25] Bleeding was defined as lesions with signs of either fresh or coagulated blood. A series of representative images of inflammation and bleeding taken by capsule endoscopy in SB were shown (Supplementary Figures 1 and 2). The detection rates of inflammation, ulcer, and polyps were similar to those in other studies.[24] However, the detection rates for bleeding and vascular diseases in our study were lower than those in other studies, particularly for OGIB. We believe that the main reason for this is that we randomly enrolled our study participants, who were either hospital inpatients or asymptomatic individuals in the general population who had physical examination and SB-CE. Because we did not specifically recruit patients with OGIB, the frequency of bleeding was expected to be much lower than the selected patient population with suspected OGIB. Our center previously published an article[26] on patients with OGIB who had negative esophagogastroduodenoscopies and colonoscopy results. In that article, the positive detection rate of bleeding was found to be 68.9%. The result was comparable to the detection rate of bleeding worldwide, as Liao et al reported in a systematic review.[27] Therefore, the reason for the lower detection rate or severity of bleeding in our study is that we randomly selected our patient population with unknown risk of OGIB.

The CNN-based auxiliary reading system we have developed is intended to be used in centers with CE equipment to serve as an auxiliary system to reduce the reading time and increase the detection rate during reviewing. The implementation of this model requires a standard graphics processing unit engine and the CNN-based auxiliary reading system, which can be integrated into the software used for reviewing CE. This study was conducted using Ankon capsule endoscopy equipment, and we will also try to verify the effectiveness of this algorithm in other types of capsule endoscopy reviewing in the future.

There are several limitations to this study. First, data for this study were collected retrospectively, and well-designed prospective studies are needed to confirm the results. Second, similar to colonoscopy studies, there is no true criterion standard for interpretation of SB-CE; therefore, there may be an underlying miss rate that we cannot assess. Third, because patients included in this study were unselected, the detection rates of some kinds of SB lesions differed from those reported in other studies. Therefore, the results may lack generalizability. Fourth, although the CNN-based auxiliary reading model can detect SB abnormalities, it cannot further classify the SB abnormalities. Therefore, further studies may be needed to develop algorithms for the classification of specific SB diseases. Finally, even though we included 5000 patients in this study, investigation in larger data sets should be conducted to verify the results before clinical use.

In summary, we present a deep-learning–based AI model for differentiating abnormal images from normal images in SB-CE examination. The CNN-based auxiliary model was then verified in a large-scale cohort, and it achieved excellent performance by providing a rapid reading time and efficient detection rate. We believe that the CNN-based auxiliary reading system proposed in our study may be an important advance for SB-CE reading and could significantly reduce the cost of SB-CE reading.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at https://doi.org/10.1053/j.gastro.2019.06.025.

References

1. Aktas H, Mensink PB. Small bowel diagnostics: current place of small bowel endoscopy. Best Pract Res Clin Gastroenterol 2012;26:209–220.
2. Flemming J, Cameron S. Small bowel capsule endoscopy: indications, results, and clinical benefit in a university environment. Medicine (Baltimore) 2018;97:e0148.
3. Rondonotti E, Spada C, Adler S, et al. Small-bowel capsule endoscopy and device-assisted enteroscopy for diagnosis and treatment of small-bowel disorders: European Society of Gastrointestinal Endoscopy (ESGE) technical review. Endoscopy 2018;50:423–446.
4. Gerson LB, Fidler JL, Cave DR, et al. ACG clinical guideline: diagnosis and management of small bowel bleeding. Am J Gastroenterol 2015;110:1265–1287.
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;89:189–194.
6. Niv Y, Niv G. Capsule endoscopy examination—preliminary review by a nurse. Dig Dis Sci 2005;50:2121–2124.
7. Costamagna G, Shah SK, Riccioni ME, et al. A prospective trial comparing small bowel radiographs and video capsule endoscopy for suspected small bowel disease. Gastroenterology 2002;123:999–1005.
8. Ribeiro E, Uhl A, Wimmer G, et al. Exploring deep learning and transfer learning for colonic polyp classification. Comput Math Methods Med 2016;2016:6584725.
9. **Esteva A, Kuprel B**, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017;542:115–118. Q16
10. Byrne MF, Chapados N, Soudan F, et al. Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. Gut 2019;68:94–100.
11. **Zhao AJ, Qian YY, Sun H**, et al. Screening for gastric cancer with magnetically controlled capsule gastroscopy in asymptomatic individuals. Gastrointest Endosc 2018;88:466–474. Q17
12. Wei W, Ge ZZ, Lu H, et al. Purgative bowel cleansing combined with simethicone improves capsule endoscopy imaging. Am J Gastroenterol 2008;103:77–82.
13. Viazis N, Sgouros S, Papaxoinis K, et al. Bowel preparation increases the diagnostic yield of capsule endoscopy: a prospective, randomized, controlled study. Gastrointest Endosc 2004;60:534–538.
14. Kaiming He XZ, Shaoqing Ren, Jian Sun. Deep residual learning for image recognition. arXiv Web site. https://arxiv.org/abs/1512.03385. Q18
15. Koulaouzidis A, Iakovidis DK, Karargyris A, et al. Optimizing lesion detection in small-bowel capsule endoscopy: from present problems to future solutions. Expert Rev Gastroenterol Hepatol 2015;9:217–235.
16. Yung DE, Sykes C, Koulaouzidis A. The validity of suspected blood indicator software in capsule endoscopy: a systematic review and meta-analysis. Expert Rev Gastroenterol Hepatol 2017;11:43–51.
17. Saurin JC, Lapalus MG, Cholet F, et al. Can we shorten the small-bowel capsule reading time with the "Quickview" image detection system? Dig Liver Dis 2012;44:477–481.
18. Saurin JC, Jacob P, Heyries L, et al. Multicenter prospective evaluation of the express view reading mode for small-bowel capsule endoscopy studies. Endosc Int Open 2018;6:E616–E621.
19. Zhou T, Han G, Li BN, et al. Quantitative analysis of patients with celiac disease by video capsule endoscopy: a deep learning method. Comput Biol Med 2017;85:1–6.

20. Buchner AM, Shahid MW, Heckman MG, et al. Trainee participation is associated with increased small adenoma detection. Gastrointest Endosc 2011;73:1223–1231.
21. 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.
22. 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. In press. https://doi.org/10.1136/gutjnl-2018-317500.
23. McAlindon ME, Ching HL, Yung D, et al. Capsule endoscopy of the small bowel. Ann Transl Med 2016;4:369.
24. Lim YJ, Lee OY, Jeen YT, et al. Indications for detection, completion, and retention rates of small bowel capsule endoscopy based on the 10-Year data from the Korean Capsule Endoscopy Registry. Clin Endosc 2015;48:399–404.
25. Zhang Y, Wu SY, Du YQ, et al. Epidemiology of obscure gastrointestinal bleeding in China: a single-center series and comprehensive analysis of literature. J Dig Dis 2018;19:33–39.
26. Gengcheng Hu JL, Xiaoping Xie, Xiaohua Hou. Application value of capsule endoscopy in the diagnosis of patients with obscure gastrointestinal bleeding. Chin J Dig Endosc 2011;28.
27. Liao Z, Gao R, Xu C, et al. Indications and detection, completion, and retention rates of small-bowel capsule endoscopy: a systematic review. Gastrointest Endosc 2010;71:280–286.

Author names in bold designate shared co-first authorship.

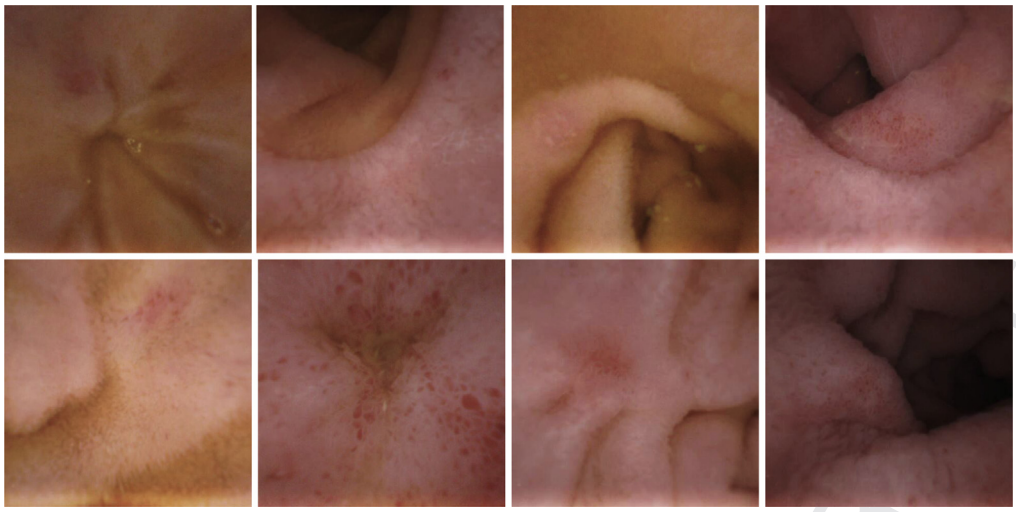
Received December 7, 2018. Accepted June 17, 2019.

Reprint requests
Address requests for reprints to: Rong Lin, MD, PhD, Department of Gastroenterology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China. e-mail: selinalin35@hotmail.com.

Acknowledgements
Author contributions: Rong Lin designed and supervised the study, including all data collection and analysis; Zhen Ding and Huiying Shi performed most of the investigation, including data collection and analysis, and wrote the manuscript; Hao Zhang and Fanhua Ming provided the technical support of the algorithm and the CNN-based AI model; Lingjun Meng, Mengke Fan, Chaoqun Han, and Kun Zhang helped collect data; Xiaoping Xie, Jun Liu, and Xiaohua Hou provide guidance in data analysis and study design; and Hao Liu was responsible for managing ESView data platform. All authors have read and approved the manuscript.

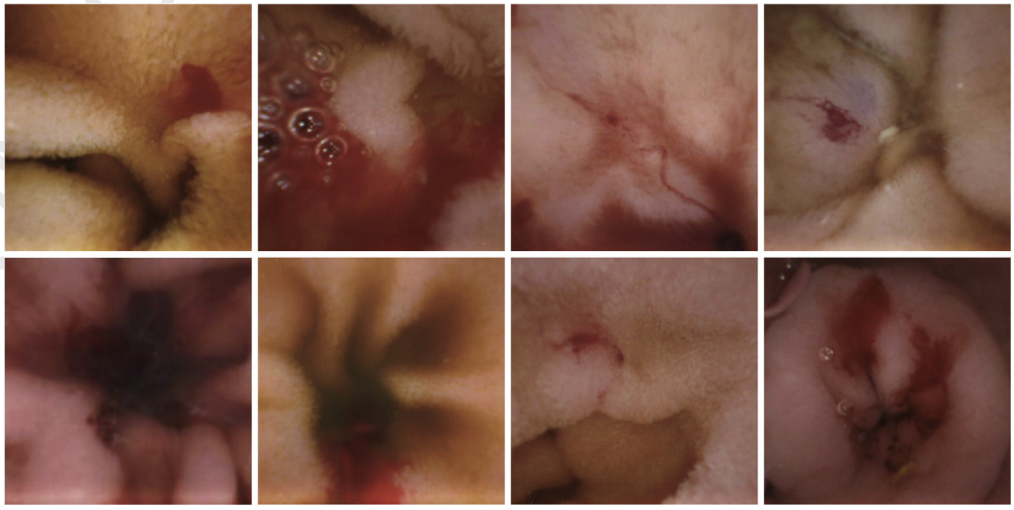
Conflicts of interest
These authors disclose the following: Hao Zhang, Fanhua Ming, and Hao Liu are research staff members of Ankon Technologies Inc. The remaining authors disclose no conflicts.

Funding
This study was supported by the National Natural Science Foundation of China (nos. 81770539, 81330014, 81572428, and 81272656), the National Key Research and Development program of China (no. 2017YFC0110003), and in part by Ankon. Wuhan Union Hospital and Ankon collaborated in this study.



print & web 4C/FPO

Supplementary Figure 1. Representative inflammation lesions pictured by CE in the SB.



print & web 4C/FPO

Supplementary Figure 2. Representative bleeding lesions pictured by CE in the SB.

Supplementary Table 1. Evaluation of the Quality of SB Cleansing

Patient no.	Clean[a] images, n	Total SB images, n	SB clean duration, *s*	SBTT, *s*	Percentage of clean time (%)	Objective score[b]
1	21,576	21,586	23,799	23,818	99.92	0.08
2	21,020	21,811	24,743	25,306	97.78	2.22
3	23,195	23,339	21,927	22,185	98.84	1.16
4	19,026	22,029	15,667	19,017	82.38	17.62
5	19,618	19,865	21,403	21,921	97.64	2.36
6	19,293	19,503	20,269	20,461	99.06	0.94
7	19,814	19,895	22,143	22,188	99.80	0.20
8	17,857	18,719	17,113	20,071	94.54	5.46
9	20,844	20,898	21,487	21,595	99.50	0.50
10	17,587	18,700	18,197	20,071	90.66	9.34

NOTE. We randomly selected 10 patients in the validation group and evaluated the quality of SB cleansing. SBTT, small-bowel transit time.
[a]The SB mucosa was defined as *clean* if, at any time, less than 25% of the mucosal surface was covered by intestinal contents or food debris.
[b]The percentage of SBTT during which the small intestinal mucosa was not clean was then calculated as an objective score. SB cleansing was considered adequate if the objective score was $<10\%$ and inadequate if $\geq 10\%$.

Supplementary Table 2. The Confusion Matrix of the CNN-Based AI Model in the Training Phase

		Predicted condition, n	
		Normal	Abnormal
True condition, n	Normal	32,052	992
	Abnormal	327	6812

Supplementary Table 3. Diagnoses of SB Abnormalities by Using the Conventional and CNN-Based Auxiliary Reading Methods in a Per-Patient Analysis

		CNN-based auxiliary reading group			
		+	+	−	−
Conventional reading group	Consensus evaluation	+	−	+	−
+	+	2443		3	
+	−		0		0
−	+	833		1	
−	−		0		1720

Supplementary Table 4. Diagnoses of SB Abnormalities by Using the Conventional and CNN-Based Auxiliary Reading Methods in a Per-Lesion Analysis

		CNN-based auxiliary reading group			
		+	+	−	−
Conventional reading group	Consensus evaluation	+	−	−	−
+	+	3230		4	
+	−		0		0
−	+	972		0	
−	−		0		1720

Supplementary Table 5. The Detection Rate for SB Abnormal Lesions or Normal Variants by the Conventional or CNN-Based Auxiliary Reading Method According to the Total Number of SB-CE Diagnoses

	Conventional reading		CNN-based auxiliary reading		
Types of SB-CE diagnoses	n	Detection rate, %	n	Detection rate, %	*P* value
SB abnormal lesions	2455	41.43	2785	47.00	
Inflammation	1576	26.59	1679	28.33	.034
Ulcer	366	6.18	372	6.28	.820
Polyps	204	3.44	261	4.40	.007
Protruding lesion	128	2.16	227	3.83	<.0001
Vascular disease	126	2.13	183	3.09	.001
Bleeding	31	0.52	39	0.66	.338
Parasite	12	0.20	12	0.20	1
Diverticulum	12	0.20	12	0.20	1
Normal variants	779	13.14	1417	23.91	
Lymphangiectasia	438	7.39	853	14.39	<.0001
Lymphatic follicular hyperplasia	120	2.02	257	4.34	<.0001
Other	221	3.73	307	5.18	<.0001

Supplementary Table 6. The Detection Rate for SB Abnormal Lesions or Normal Variants by the Conventional or CNN-Based Auxiliary Reading Method According to the Total Number of Patients

	Conventional reading		CNN-based auxiliary reading		
Types of SB-CE diagnoses	n	Detection rate, %	n	Detection rate, %	*P* value
SB abnormal lesions					
Inflammation	1576	31.52	1679	33.58	.028
Ulcer	366	7.32	372	7.44	.818
Polyps	204	4.08	261	5.22	.007
Protruding lesion	128	2.56	227	4.54	<.0001
Vascular disease	126	2.52	183	3.66	.001
Bleeding	31	0.62	39	0.78	.337
Parasite	12	0.24	12	0.24	1
Diverticulum	12	0.24	12	0.24	1
Normal variants					
Lymphangiectasia	438	8.76	853	17.06	<.0001
Lymphatic follicular hyperplasia	120	2.40	257	5.14	<.0001
Other	221	4.42	307	6.14	<.0001

Supplementary Table 7. Interobserver Agreement Among 20 Gastroenterologists in Their Capsule Interpretations in Per-Patient and Per-Lesion Analyses

Type of analysis	Conventional reading group, κ	CNN-based auxiliary reading group, κ
Per patient	0.811	0.937
Per lesion	0.783	0.944