ORIGINAL ARTICLE

Use of artificial intelligence for detection of gastric lesions by magnetically controlled capsule endoscopy

Ji Xia, MD,[1,*] Tian Xia, MD,[1,*] Jun Pan, MD,[1,*] Fei Gao, MD, PhD,[2,*] Shuang Wang, MM,[2] Yang-Yang Qian, MD,[1] Heng Wang, PhD,[2] Jie Zhao, MS,[2] Xi Jiang, MD,[1] Wen-Bin Zou, MD,[1] Yuan-Chen Wang, MD,[1] Wei Zhou, MD,[1] Zhao-Shen Li, MD,[1] Zhuan Liao, MD[1]

Shanghai, Beijing, China

Background and Aims: Magnetically controlled capsule endoscopy (MCE) has become an efficient diagnostic modality for gastric diseases. We developed a novel automatic gastric lesion detection system to assist in diagnosis and reduce inter-physician variations. This study aimed to evaluate the diagnostic capability of the computer-aided detection system for MCE images.

Methods: We developed a novel automatic gastric lesion detection system based on a convolutional neural network (CNN) and faster region-based convolutional neural network (RCNN). A total of 1,023,955 MCE images from 797 patients were used to train and test the system. These images were divided into 7 categories (erosion, polyp, ulcer, submucosal tumor, xanthoma, normal mucosa, and invalid images). The primary endpoint was the sensitivity of the system.

Results: The system detected gastric focal lesions with 96.2% sensitivity (95% confidence interval (CI), 95.7%-96.5%), 76.2% specificity (95% CI, 75.97%-76.3%), 16.0% positive predictive value (95% CI, 15.7%-16.3%), 99.7% negative predictive value (95% CI, 99.74%-99.79%), and 77.1% accuracy (95% CI, 76.9%-77.3%) (sensitivity was 99.3% for erosions; 96.5% for polyps; 89.3% for ulcers; 87.2% for submucosal tumors; 90.6% for xanthomas; 67.8% for normal; and 96.1% for invalid images). Analysis of the receiver operating characteristic curve showed that the area under the curve for all positive images was 0.84. Image processing time was 44 milliseconds per image for the system and 0.38 ± 0.29 seconds per image for clinicians ($P < .001$). The kappa value of 2 times repeated reads was 1.

Conclusions: The CNN faster-RCNN-based diagnostic program system showed good performance in diagnosing gastric focal lesions in MCE images. (Gastrointest Endosc 2020;■:1-7.)

INTRODUCTION

The magnetically controlled capsule endoscopy (MCE) system has become a noninvasive diagnostic modality for gastric diseases in recent years.[1-3] In a multicenter study enrolling 350 patients, MCE demonstrated 90.4% sensitivity and 94.7% specificity in detecting focal lesions in the stomach.[2] MCE also showed potential for gastric cancer screening in the general population.[3] However, inter-physician variations, such as differences in the skill of the operators, perceptual factors, personality characteristics, knowledge, and attitudes, were

Abbreviations: AI, artificial intelligence; AUC, area under curve; CAD, computer-aided diagnosis; CI, confidence intervals; CNN, convolutional neural networks; MCE, magnetically controlled capsule endoscopy; NPV, negative predictive value; PPV, positive predictive value; RCNN, region-based convolutional neural network.

DISCLOSURE: *All authors disclosed no financial relationships.*

*Drs J. Xia, T. Xia, Pan, and Gao contributed equally to this article.

0016-5107/$36.00
https://doi.org/10.1016/j.gie.2020.05.027

Received February 9, 2020. Accepted May 1, 2020.

Current affiliations: National Clinical Research Center for Digestive Diseases, Department of Gastroenterology, Changhai Hospital, Second Military Medical University, Shanghai (1); Beijing Medicinovo Technology Co. Ltd., Beijing, China (2).

Reprint requests: Zhuan Liao and Zhao-Shen Li, National Clinical Research Center for Digestive Diseases, Department of Gastroenterology, Changhai Hospital, Second Military Medical University, 168 Changhai Road, Shanghai 200433, China.

If you would like to chat with an author of this article, you may contact Dr Liao at zhuanliao@hotmail.com or Dr Li at zhaoshenli@smmu.edu.cn.

found in GI endoscopy examinations,[4] including MCE examinations.

Recently, artificial intelligence (AI), especially deep learning, has dominated the field of image recognition, relying on increasingly powerful computing power and big data.[5,6] Various types of deep convolutional neural networks (CNN) have been developed and gradually applied to various types of visual tasks, such as classification, detection, and segmentation.[7-10] Liu et al[11] developed a framework for automated detection and localization of tumors as small as 100 × 100 pixels in gigapixel microscopy images of 100,000 × 100,000 pixels[11] and detected 92.4% of the tumors with 8 false positives per image. Esteva et al[12] used a single CNN for classification of skin lesions and achieved performances equal to that of all experts tested. Ding et al[13] developed an artificial intelligence algorithm to assist physicians in detecting 99.9% lesions in the small bowel successfully. Wu et al[14] even constructed a real-time quality improving system to reduce the blind spot rate from 22.5% to 5.9% in EGD. Therefore, a novel automatic lesion detection system could eliminate the variation between physicians, thereby increasing the stability of screening results.

So far, no computer-aided diagnostic system for capsule endoscopy in gastric examinations has been reported. In this study, we aimed to develop an AI-based auxiliary diagnostic system based on deep learning that could identify images with 5 common focal lesions (erosion, polyp, ulcer, submucosal tumor, and xanthoma) and test the performance of the system.

METHODS

Materials

The NaviCam MCE system (Ankon Technologies, China) has been described in detail previously.[2,15] In brief, it consists of a robotic magnetic arm, a workstation, and a capsule endoscope. During examination, physicians manipulate the joysticks on the workstation to mobilize the robotic magnetic arm to precisely control the movement of the capsule. NaviCam had a sensitivity of 90.4% and a specificity of 94.7% for diagnosing gastric lesions[2] and outperformed EGD in terms of comfort, no need for sedation, and no risk of cross-infection (single use).[16,17] For gastric lesions detected by MCE that required a biopsy, EGD with biopsy was performed subsequently to establish the final diagnosis.

The model developed in this study was based on supervised learning. We collected a total of 822,590 images pertaining to 697 consecutive patients undergoing MCE examinations in Changhai Hospital from July 2014 to December 2017 to train the model. We labeled all images into 7 categories: erosion, polyp, ulcer, submucosal tumor, xanthoma, normal, and invalid (not pertaining to the stomach or very poor image quality). Gastric cancer was not included because the number of images from the patients (n = 4) was too small to train the model. To prevent over-fitting of the model, we used data augmentation techniques, including random rotation, random shifting, and random zooming, which guaranteed that the model would not encounter exactly the same picture again throughout the training process.

System construction

Because it was difficult to guarantee both high sensitivity and high specificity using a single CNN, we designed our system with 3 CNNs to form a pipeline. The first CNN used Resnet-34 architecture, which categorized 3 types of images: images with lesions, images without lesions (normal), and invalid images. The form of the output was the probability that an input image belonged to these 3 types. The second CNN also used Resnet-34 architecture, which categorized 5 types of lesions (erosion, polyp, ulcer, submucosal tumor, and xanthoma) and an additional type, normal. The first 2 CNNs were classification models and good at dealing with large lesions. For dealing with multiple small lesions, we used a detection model, faster-region-based convolutional neural network (RCNN), in parallel, which focused on preventing the small lesions from being missed and ensuring that the system detect multi-targets. A schematic illustration of the 3 CNNs forming a pipeline is shown in Figure 1. The 6 types of lesions (erosion, polyp, ulcer, submucosal tumor, xanthoma, and normal) were assessed by 2 serial CNNs (Net1 + Net2 or Net1 + Net3). The first one guaranteed high sensitivity and the second one guaranteed high specificity.

Test method

To test the performance of the system, we collected 201,365 images pertaining to 100 consecutive patients undergoing MCE examinations in Changhai Hospital from January 2018 to May 2018 as the test set. The output of the system was an array of the probability that an input image belonged to one of the 7 categories. To assess the ability of the system to detect lesions, the categories of the 5 lesions were defined as positive, and the remaining 2 categories were defined as negative. The diagnosis result of the system was the category with the highest probability. Diagnosis by 2 experienced clinicians (T.X. and J.P.) with specialization in gastroenterology was regarded as the criterion standard, and discrepancies were resolved by consensus (Supplementary Table 1, available online at www.giejournal.org). The system ran on a mainframe with dimensions 450 × 185 × 420 mm (GPU with 1480 MHz frequency and 11 GB video memory; CPU with 3.7 GHz frequency and 32 GB memory).

Statistical analysis

For the sample size calculation, considering the results of the experts as the criterion standard, our study assumed that the detection system had at least 90% sensitivity and 75% specificity in detecting images with gastric focal lesions.

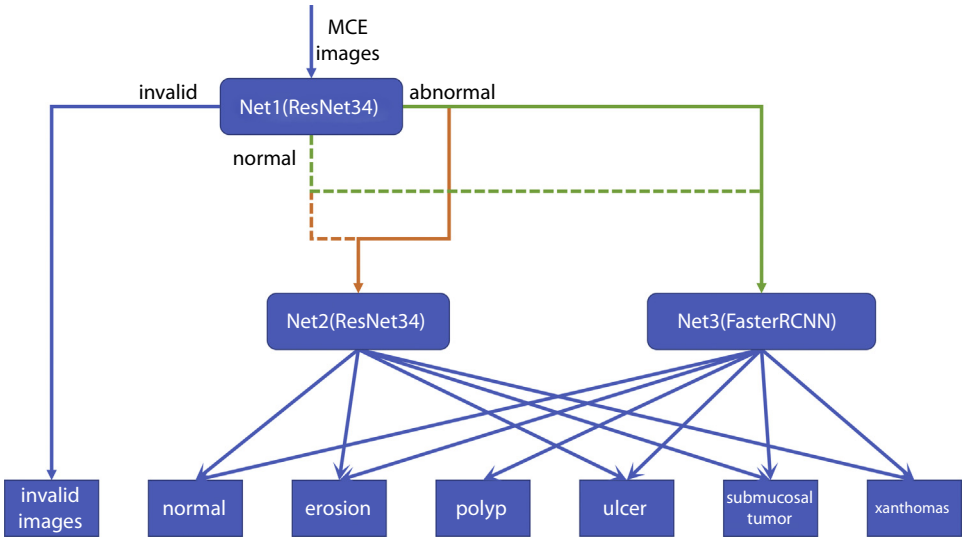


Figure 1. Schematic illustration of the entire architecture of the system consisting of 3 convolutional neural networks. *MCE*, Magnetically controlled capsule endoscopy; *RCNN*, region-based convolutional neural network.

TABLE 1. Classification of images by 2 clinicians as the standard to evaluate the diagnostic ability of the system for the test dataset (100 patients)

Classification	Number of images	Number of lesions
Positive		
Erosion	5060	42
Polyp	1754	17
Ulcer	975	9
Submucosal tumor	743	5
Xanthoma	543	11
Negative		
Normal	135,253	–
Invalid images	57,037	–
Total	201,365	84

To maintain that hypothesis with a significance level of 5% (2-sided) and tolerance error of 5%, the 158 positive findings and 306 negative findings were required. A total study size of at least 612 MCE images was required. Sample size was calculated by PASS 11 using objective performance criteria.

Receiver operating characteristic curves were plotted for classification of each category, and the area under the curves (AUCs) were calculated by scikit-learn machine learning in Python 0.20.3. The sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, and 95% confidence intervals (CI) were calculated using SAS 9.4. Mann-Whitney U test was used to compare the image processing time between the system and clinicians using SPSS 22.0; P values <.05 were considered indicative of statistical significance.

RESULTS

Test dataset

In the test dataset (Table 1), 100 patients contributed to 84 focal lesions (including erosion, 42; polyp, 17; ulcer, 9; submucosal tumor, 5; xanthoma, 11) and 201,365 MCE images, including 9075 positive images (erosion, 5060; polyp, 1754; ulcer, 975; submucosal tumor, 743; xanthoma, 543) and 192,290 negative images (normal, 135,253; invalid images, 57,037).

Performance of the system

The system detected positive gastric lesion images with 96.2% sensitivity (95% CI, 95.7%-96.5%), 76.2% specificity (95% CI, 75.97%-76.35%), 16.0% PPV (95% CI, 15.7%-16.3%), 99.7% NPV (95% CI, 99.7%-99.79%), and 77.1% accuracy (95% CI, 76.9%-77.2%) (Table 2). The AUC for all positive images was 0.84 (Fig. 2). For per-lesion analysis, the sensitivity, specificity, PPV, NPV, and accuracy for detecting gastric lesions were 100% (95% CI, 95.7%-100%), 56.5% (95% CI, 44.0%-68.4%), 73.7% (95% CI, 64.6%-81.5%), 100% (95% CI, 91.0%-100%), and 80.4% (95% CI, 73.2%-86.4%), respectively. For per-patient analysis, the sensitivity, specificity, PPV, NPV, and accuracy for diagnosing patients with gastric lesions were 100% (95% CI, 92.5%-100%), 73.6% (95% CI, 59.7%-89.7%), 77.0% (95% CI, 64.5%-86.9%), 100% (95% CI, 91.0%-100%), and 86.0% (95% CI, 77.6%-92.1%), respectively (Table 2).

The sensitivity of the system for recognizing the image of each category was as follows: 99.3% (95% CI, 98.9%-99.5%) for erosion; 96.5% (95% CI, 95.5%-97.3%) for polyp; 89.3% (95% CI, 87.2%-91.2%) for ulcer; 87.2% (95% CI, 84.6%-89.5%) for submucosal tumor; 90.6% (95% CI, 87.9%-92.9%) for xanthoma; 67.8% (95% CI, 67.5%-70.0%) for normal; and 96.1% (95% CI, 96.0%-96.3%) for invalid images (Table 3).

TABLE 2. Diagnostic performance of the system for gastric lesions in per-image, per-lesion, and per-patient analyses

	Per image analysis	Per lesion analysis	Per patient analysis
Sensitivity (%), 95% CI	96.2 (95.7-96.5)	100 (95.7-100)	100 (92.5-100)
Specificity (%), 95% CI	76.2 (75.97-76.3)	56.5 (44.0-68.4)	73.6 (59.7-89.7)
PPV (%), 95% CI	16.0 (15.7-16.3)	73.7 (64.6-81.5)	77.0 (64.5-86.9)
NPV (%), 95% CI	99.7 (99.7-99.8)	100 (91.0-100)	100 (91.0-100)
Accuracy (%), 95% CI	77.1 (76.9-77.3)	80.4 (73.2-86.4)	86.0 (77.6-92.1)

CI, Confidence interval; *PPV*, positive predictive value; *NPV*, negative predictive value.

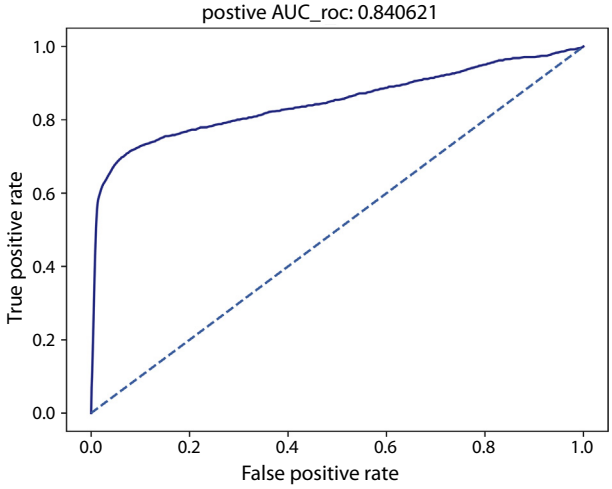


Figure 2. Area under receiver operating characteristic (AUC-roc) curve for positive images of gastric lesions.

The AUC values for each category were as follows: erosion, 0.90 (95% CI, 0.85-0.95); polyp, 0.898 (95% CI, 0.84-96); ulcer, 0.87 (95% CI, 0.84-0.90); submucosal tumor, 0.88 (95% CI, 0.81-0.96); xanthoma, 0.90 (95% CI, 0.87-0.93); normal, 0.80 (95% CI, 0.69-0.92); and invalid images, 0.902 (95% CI, 0.82-0.98) (Supplementary Fig. 1, available online at www.giejournal.org). Representative examples of images screened by 3 models of the system are shown in Supplementary Figure 2 (available online at www.giejournal.org) and Figures 3 and 4.

Image processing time and reproducibility

The average processing time of the CNN faster RCNN-based diagnostic system was 44 ± 16 milliseconds per image, and the average processing time of the clinicians was 0.38 ± 0.29 seconds per image. There was a significant difference between the processing time of the system and the clinicians ($P < .001$). The kappa value of 2 times repeated reads was 1. The reproducibility of the system was perfect.

DISCUSSION

In recent years, AI has shown its power in the field of GI endoscopy image recognition due to its powerful image processing capabilities. In EGD, researchers have developed a variety of computer-aided diagnosis (CAD) systems for recognition of anatomic locations, diagnosis of esophageal cancer, diagnosis of *Helicobacter pylori* infection, detection of gastric cancers, and diagnosis of depth of invasion in gastric cancer to assist physicians in diagnosis.[17-22] With regard to upper GI cancers, Luo et al[23] demonstrated that a real-time AI detection system by EGD had a diagnostic accuracy of 92.7%. Due to technical limitations, biopsy of lesions cannot be performed by MCE yet. Its function aims at population screening and detection of suspicious lesions, which requires a detection system for multi-classification of lesions. However, no CAD system has been designed to assist physicians in performing gastric capsule endoscopy examinations such as MCE.

In capsule endoscopy, especially small-bowel examination, scientists have developed CAD systems to detect diseases, including small intestinal polyps, erosions, ulcers, celiac disease, inflammatory bowel disease, and bleeding, to reduce tedious review time and increase the accuracy of medical examinations.[13,24-28] The CNN-based auxiliary reading model developed by Ding et al[13] based on more than 100 million images identified small-bowel abnormalities with sensitivities of 99.88% and 99.90% in per patient and per lesion analyses, respectively, and reading time was significantly reduced compared with conventional reading. Due to the complex gastric environment and similar features between lesions, no multi-classification diagnostic system for gastric lesions has been developed to date.

In this study, we explored a CNN faster-RCNN-based diagnostic system to facilitate evaluation of MCE images, and successfully validated the model using 201,365 images. Several novel aspects of our study should be emphasized. First, to the best of our knowledge, this is the first AI diagnostic system to evaluate gastric images of capsule endoscopy used in clinical practice. Second, this is the first report that describes the development of a CNN-based deep learning model to facilitate the diagnosis of a wide variety of abnormal gastric lesions, including erosion, polyp, ulcer, submucosal tumor, and xanthoma. Third, we built a CNN-based algorithm with a high sensitivity (ranging from 87.2% to 99.3%) and high time efficiency (processing time per image of 44 ± 16 milliseconds vs 0.38 ± 0.29 seconds for clinicians) for detection of gastric lesions using MCE.

TABLE 3. Diagnostic performance of the system detection of erosion, polyp, ulcer, submucosal tumor, xanthoma, normal, and invalid images in the per image analysis

Lesions	Sensitivity (%), (95% CI)	Specificity (%), (95% CI)	PPV (%), (95% CI)	NPV (%), (95% CI)	Accuracy (%), (95% CI)
Erosion	99.3 (98.99-99.5)	96.8 (96.7-96.9)	44.3 (43.3-45.2)	99.98 (99.9-99.9)	96.8 (96.8-96.9)
Polyp	96.5 (95.5-97.3)	94.8 (94.7-94.9)	14.1 (13.5-14.7)	99.97 (99.9-99.9)	94.9 (94.8-94.95)
Ulcer	89.3 (87.2-91.2)	93.7 (93.6-93.8)	6.5 (6.1-6.9)	99.94 (99.9-99.9)	93.7 (93.6-93.8)
Submucosal tumor	87.2 (84.6-89.5)	95.3 (95.2-95.4)	6.38 (5.9-6.9)	99.95 (99.9-99.9)	95.2 (95.1-95.3)
Xanthoma	90.6 (87.9-92.9)	96.9 (96.9-97.0)	7.38 (6.8-8.0)	99.97 (99.9-99.9)	96.9 (96.8-96.98)
Normal	67.7 (67.5-70.0)	98.1 (98.0-98.2)	98.7 (98.6-98.8)	59.7 (59.4-60.0)	77.7 (77.5-77.9)
Invalid Image	96.1 (96.0-96.3)	99.95 (99.9-99.96)	99.9 (99.8-99.9)	98.5 (98.4-98.6)	98.9 (98.8-98.9)

CI, Confidence interval; *PPV*, positive predictive value; *NPV*, negative predictive value.

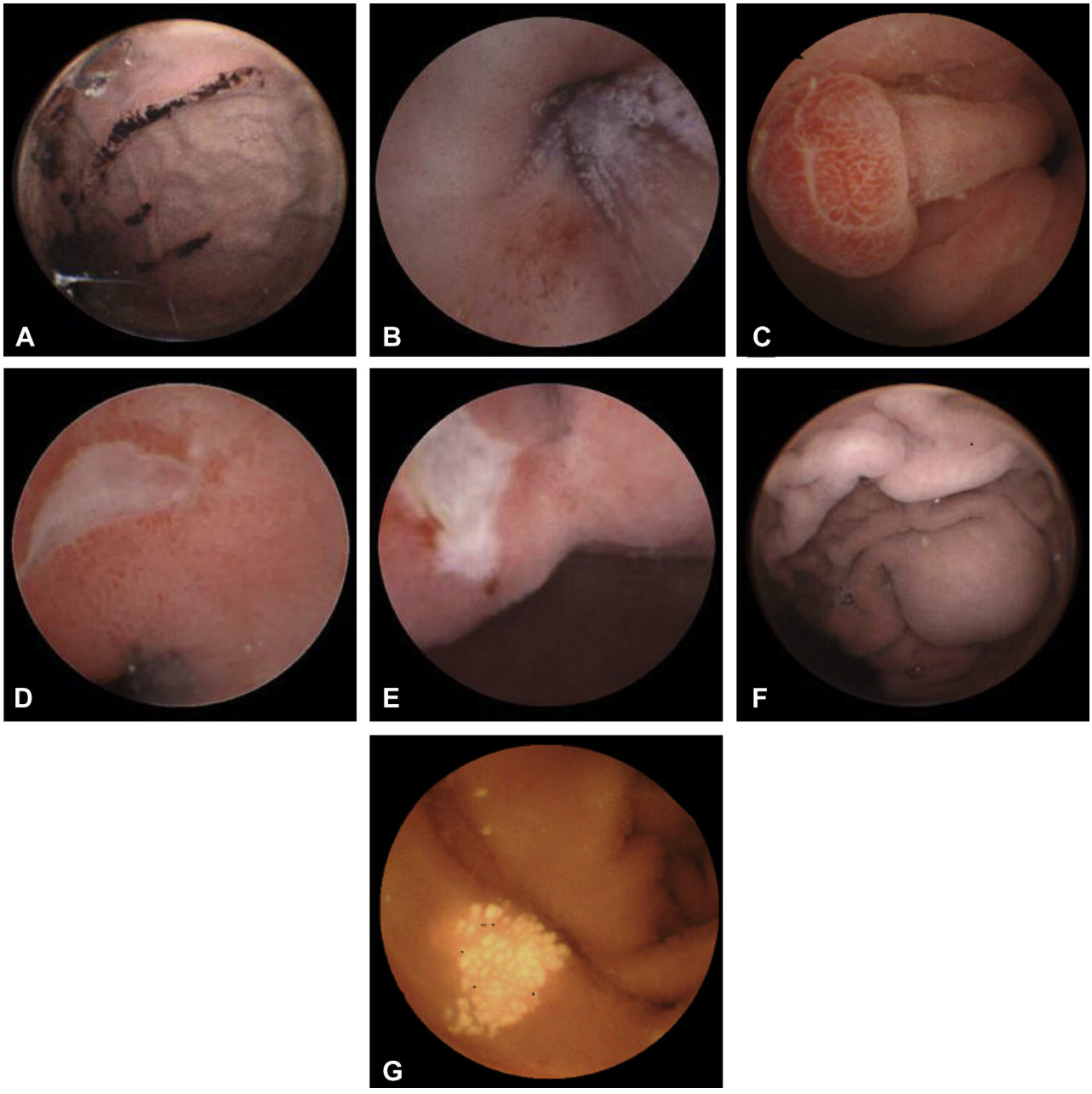


Figure 3. The output of the second CNN. The second CNN mainly categorized 5 types of lesions: (**A and B**) erosion; (**C**) polyp; (**D and E**) ulcer; (**F**) submucosal tumor; (**G**) xanthoma. *CNN*, Convolutional neural network.

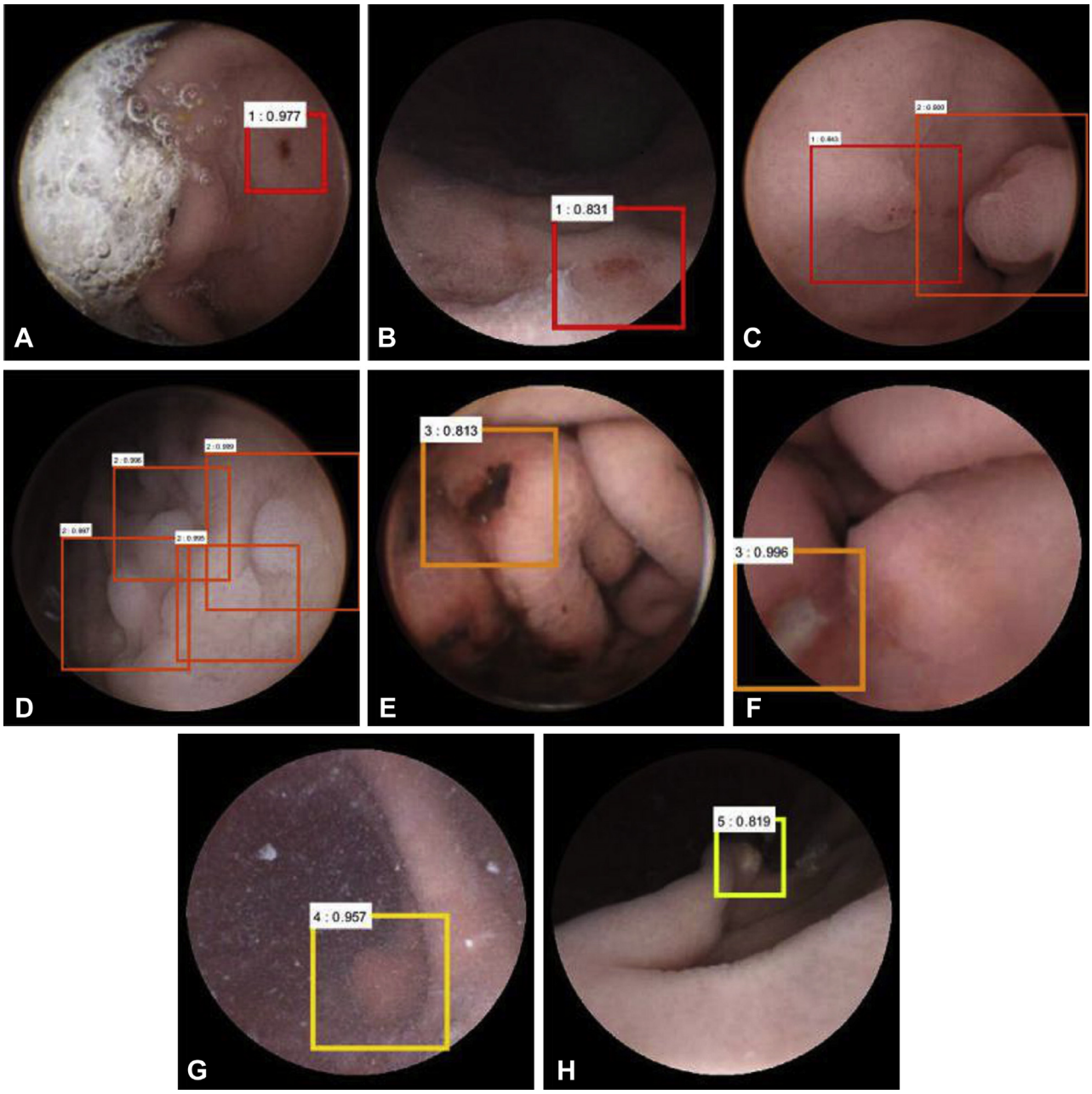


Figure 4. The output of the faster-RCNN. The faster-RCNN mainly dealt with multiple small lesions in an image. Different colored boxes and numbers represent different categories and the following decimal represented the probability: (**A and B**) erosion; (**C and D**) polyp; (**E and F**) ulcer; (**G**) submucosal tumor; (**H**) xanthoma. *RCNN*, region-based convolutional neural network.

The diagnostic performance showed potential for clinical application.

Without neglecting important lesions, the system will greatly reduce clinicians' workload; the time required to review selected frames is likely to be shorter than the total time required to analyze the entire image set. In addition, with a stable algorithm framework, there is no subjective bias in the image recognition process. The recognition result is stable and does not change with time and operating factors, which can provide a reliable reference for physicians to reduce variation. The CNN faster-RCNN-based diagnostic system showed a high overall rate of lesion detection and accuracy, laying the foundation for future mass screening of the GI tract, with remote reading, physician training, and quality control. If we reinforce the learning the parameters iteratively, the accuracy can be improved further to assist clinicians in their clinical practice. MCE has gradually evolved into a real-time examination modality. In the future, we will further improve the system's computing ability to achieve real-time detection.

Several limitations of this study should be considered when interpreting our results. First, the PPV of the system

was only 16.0%, due to the high proportion of negative images in the test dataset (192,290 negative images compared with 9075 positive images) and relatively low specificity of 76.2%, which greatly increased the false-positive data. An increase in the number of positive images will help correct the effects of excess negative data, leading to more clinically significant statistical results. Second, the study did not include images of gastric cancer due to the limited number of patients (n = 4) available for training. Future studies with targeted enrollment and multicenter collaboration should be directed to resolve this issue of paramount significance. Third, as an experimental evaluation and first-step investigation, the system was developed and tested on still images. Future studies evaluating the real-time use of AI in MCE are warranted.

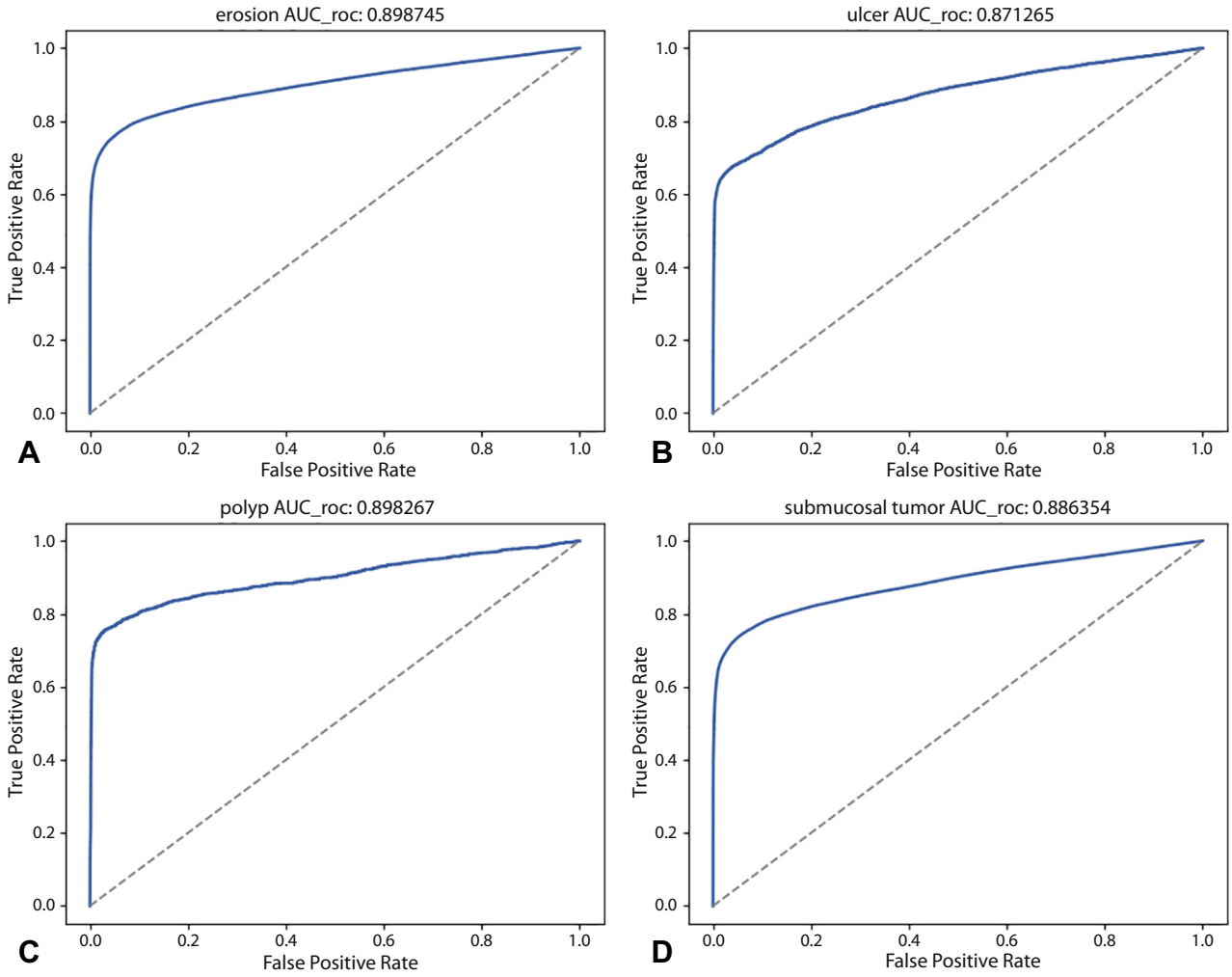
In conclusion, our trained CNN faster-RCNN-based diagnostic program showed a promising performance in the classification of MCE images, highlighting its potential for future application as an AI diagnostic system.

ACKNOWLEDGMENTS

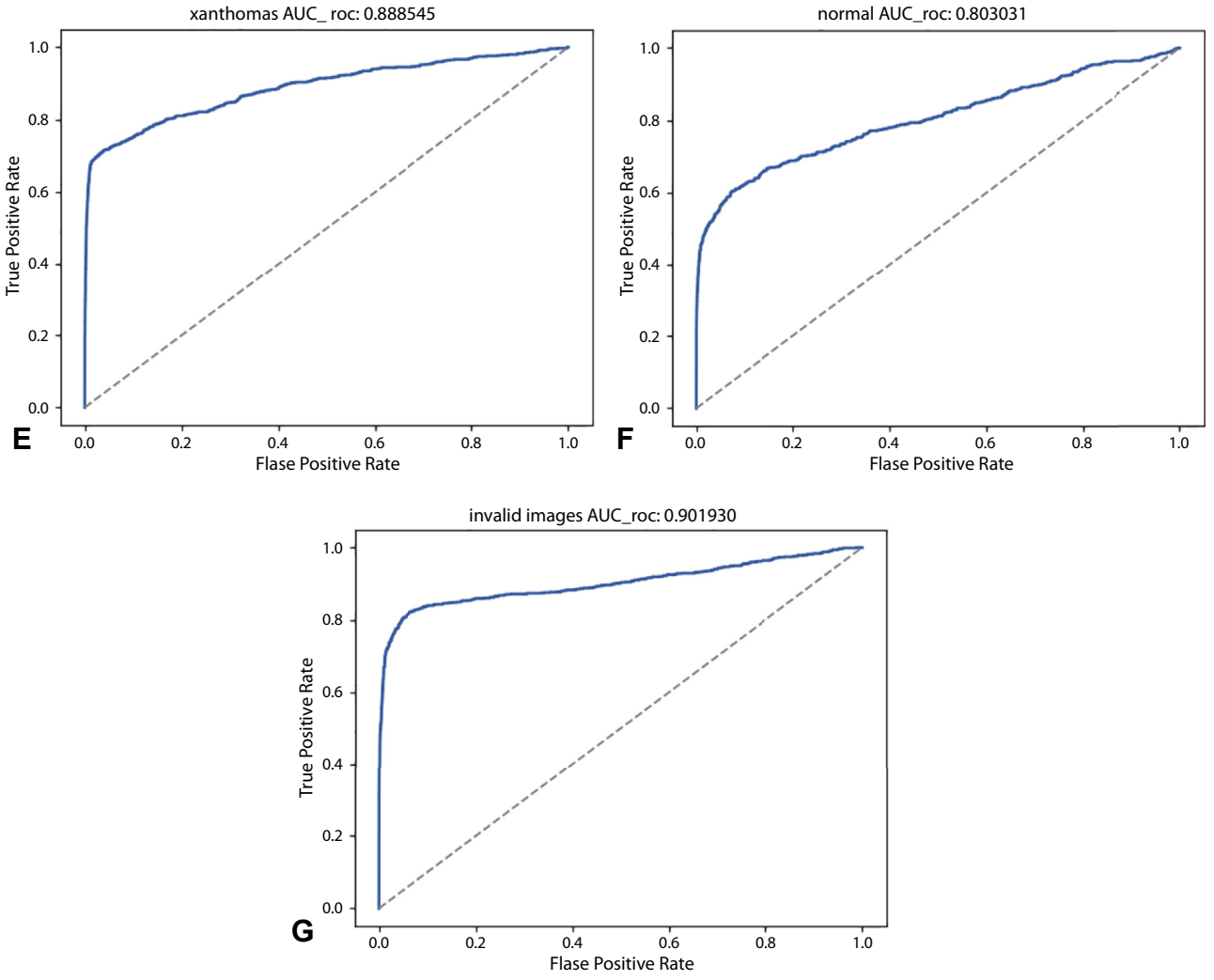
This study is supported by grants from the National Natural Science Foundation of China (to Z.L., no. 81422010); the "Ten Thousand Plan"-National High Level Talents Special Support Plan (to Z.L.); the Shuguang Program of Shanghai Education Development Foundation and Shanghai Municipal Education Commission (to Z.L., no. 15SG33); the Chang Jiang Scholars Program of Ministry of Education (to Z.L., no. Q2015190); Shanghai "Rising Stars of Medical Talent" Youth Development Program (to T.X., no.[2019]72) and Shanghai Sailing Program (to J.P., no. 18YF1422800; to Y.-Y.Q., no. 19YF1446700), China. The study funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. The opinions, results, and conclusions reported in this article are those of the authors and are independent from the funding sources.

REFERENCES

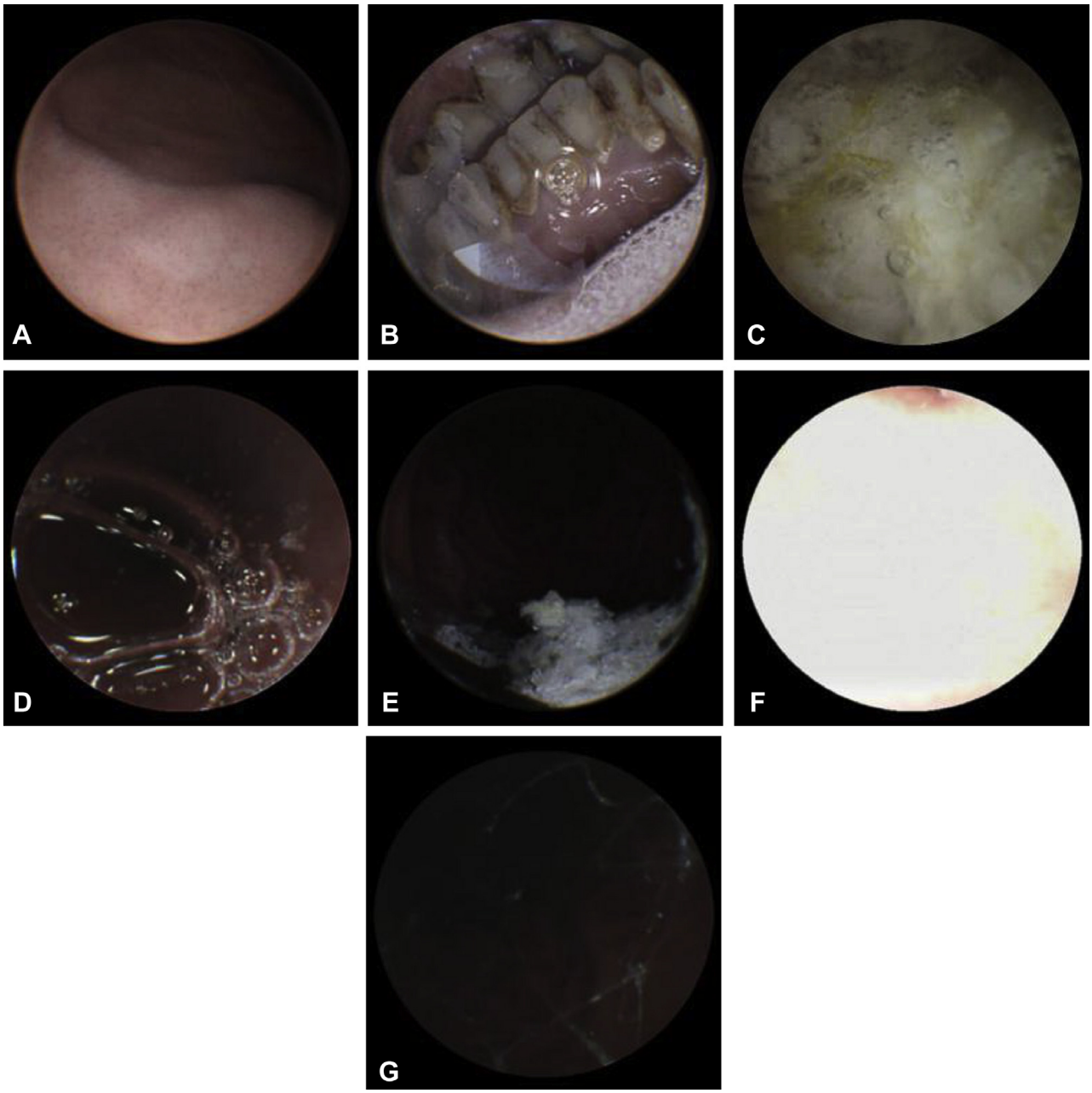
1. Liao Z, Duan XD, Xin L, et al. Feasibility and safety of magnetic-controlled capsule endoscopy system in examination of human stomach: a pilot study in healthy volunteers. J Interv Gastroenterol 2012;2:155-60.
2. Liao Z, Hou X, Lin-Hu EQ, et al. Accuracy of magnetically controlled capsule endoscopy, compared with conventional gastroscopy, in detection of gastric diseases. Clin Gastroenterol Hepatol 2016;14:1266-73.e1.
3. Zhao AJ, Qian YY, Sun H, et al. Screening for gastric cancer with magnetically controlled capsule gastroscopy in asymptomatic individuals. Gastrointest Endosc 2018;88:466-74.e1.
4. Hewett DG, Kahi CJ, Rex DK. Efficacy and effectiveness of colonoscopy: how do we bridge the gap. Gastrointest Endosc Clin N Am 2010;20:673-84.
5. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436-44.
6. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. Med Image Anal 2017;42:60-88.
7. Zhou M, Bao G, Geng Y, et al. Polyp detection and radius measurement in small intestine using video capsule endoscopy. 7th International Conference on Biomedical Engineering and Informatics, Dalian, 2014. p. 237-41.
8. Nawarathna R, Oh J, Muthukudage J, et al. Abnormal image detection in endoscopy videos using a filter bank and local binary patterns. Neurocomputing 2014;144:70-91.
9. Barbosa DC, Roupar DB, Ramos JC, et al. Automatic small bowel tumor diagnosis by using multi-scale wavelet-based analysis in wireless capsule endoscopy images. Biomed Eng Online 2012;11:3.
10. Shin HC, Roth HR, Gao M, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans Med Imaging 2016;35:1285-98.
11. Liu Y, Gadepalli K, Norouzi M, et al. Detecting cancer metastases on gigapixel pathology images. arXiv 2017;1703.02442v2.
12. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017;542:115-8.
13. 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.e5.
14. Wu L, Zhang J, Zhou W, et al. Randomised controlled trial of WISENSE, a real-time quality improving system for monitoring blind spots during esophagogastroduodenoscopy. Gut 2019;68:2161-9.
15. Jiang X, Pan J, Li ZS, et al. Standardized examination procedure of magnetically controlled capsule endoscopy. VideoGIE 2019;4:239-43.
16. Tai F, Ching HL, Hale MF, et al. Upper gastrointestinal endoscopy: can we cut the cord. Lancet Gastroenterol Hepatol 2019;4:749-51.
17. Zou WB, Hou XH, Xin L, et al. Magnetic-controlled capsule endoscopy vs. gastroscopy for gastric diseases: a two-center self-controlled comparative trial. Endoscopy 2015;47:525-8.
18. Shichijo S, Nomura S, Aoyama K, et al. Application of convolutional neural networks in the diagnosis of helicobacter pylori infection based on endoscopic images. EBioMedicine 2017;25:106-11.
19. Hirasawa T, Aoyama K, Tanimoto T, et al. Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images. Gastric Cancer 2018;21:653-60.
20. Zhu Y, Wang QC, Xu MD, et al. Application of convolutional neural network in the diagnosis of the invasion depth of gastric cancer based on conventional endoscopy. Gastrointest Endosc 2019;89:806-15.e1.
21. Yoon HJ, Kim S, Kim JH, et al. A lesion-based convolutional neural network improves endoscopic detection and depth prediction of early gastric cancer. J Clin Med 2019;8:1310.
22. Guimarães P, Keller A, Fehlmann T, et al. Deep-learning based detection of gastric precancerous conditions. Gut 2020;69:4-6.
23. Luo H, Xu G, Li C, et al. Real-time artificial intelligence for detection of upper gastrointestinal cancer by endoscopy: a multicentre, case-control, diagnostic study. Lancet Oncol 2019;20:1645-54.
24. Qiao P, Liu H, Yan X, et al. A smart capsule system for automated detection of intestinal bleeding using HSL color recognition. PLoS One 2016;11:e0166488.
25. 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.
26. 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.
27. Billah M, Waheed S, Rahman MM. An automatic gastrointestinal polyp detection system in video endoscopy using fusion of color wavelet and convolutional neural network features. Int J Biomed Imaging 2017;2017:9545920.
28. Alizadeh M, Maghsoudi OH, Sharzehi K, et al. Detection of small bowel tumor in wireless capsule endoscopy images using an adaptive neuro-fuzzy inference system. J Biomed Res 2017;31:419-27.



Supplementary Figure 1. Area under receiver operating characteristic curves (AUC-roc) for each category and all positive images. The area under the curve for each category: (**A**) erosion; (**B**) ulcer; (**C**) polyp; (**D**) submucosal tumor; (**E**) xanthomas; (**F**) normal; (**G**) invalid images.



Supplementary Figure 1. Continued.



Supplementary Figure 2. The output of the first CNN. The first CNN categorizes 3 types of scenarios: with lesions, without lesions, and invalid images (non-gastric image or poor image quality). (**A**) Normal gastric mucosa; (**B**) oral images; (**C**) gastric contents; (**D**) bubbles; (**E**) underexposure; (**F**) overexposure; (**G**) mucus. *CNN*, Convolutional neural network.

SUPPLEMENTARY TABLE 1. Diagnostic results of the 2 experts for the training and test datasets

	Expert 1						
Expert 2	**Erosion**	**Polyp**	**Ulcer**	**Submucosal tumor**	**Xanthoma**	**Normal**	**Invalid image**
Erosion	22776	–	385	–	115	455	–
Polyp	–	8175	–	237	–	–	–
Ulcer	191	–	4902	–	12	–	–
Submucosal tumor	–	–	–	3758	–	–	–
Xanthoma	–	–	45	–	3425	–	–
Normal	859	–		–	–	686,319	1168
Invalid image	–	–	–	–	–	1731	289,402

Kappa value = 0.989.