Identifying early gastric cancer under magnifying narrow-band images with deep learning: a multicenter study

Hao Hu, MD,[1,*] Lixin Gong, MS,[2,3,*] Di Dong, PhD,[3,4,*] Liang Zhu, MD,[1] Min Wang, MD,[5] Jie He, MD,[6] Lei Shu, MD,[7] Yiling Cai, MD,[8] Shilun Cai, MD,[1] Wei Su, MD,[1] Yunshi Zhong, MD,[1] Cong Li, BS,[3,4] Yongbei Zhu, MS,[3,9] Mengjie Fang, MS,[3,4] Lianzhen Zhong, BS,[3,4] Xin Yang, PhD,[3,4] Pinghong Zhou, MD,[1] Jie Tian, PhD[3,9]

Shanghai, Shenyang, Beijing, Xiamen, Wuhan, Zhangzhou, China

Background and Aims: Narrow-band imaging with magnifying endoscopy (ME-NBI) has shown advantages in the diagnosis of early gastric cancer (EGC). However, proficiency in diagnostic algorithms requires substantial expertise and experience. In this study, we aimed to develop a computer-aided diagnostic model for EGM (EGCM) to analyze and assist in the diagnosis of EGC under ME-NBI.

Methods: A total of 1777 ME-NBI images from 295 cases were collected from 3 centers. These cases were randomly divided into a training cohort (n = 170), an internal test cohort (ITC, n = 73), and an external test cohort (ETC, n = 52). EGCM based on VGG-19 architecture (Visual Geometry Group [VGG], Oxford University, Oxford, UK) with a single fully connected 2-classification layer was developed through fine-tuning and validated on all cohorts. Furthermore, we compared the model with 8 endoscopists with varying experience. Primary comparison measures included accuracy, area under the receiver operating characteristic curve (AUC), sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV).

Results: EGCM acquired AUCs of .808 in the ITC and .813 in the ETC. Moreover, EGCM achieved similar predictive performance as the senior endoscopists (accuracy: .770 vs .755, $P = .355$; sensitivity: .792 vs .767, $P = .183$; specificity: .745 vs .742, $P = .931$) but better than the junior endoscopists (accuracy: .770 vs .728, $P < .05$). After referring to the results of EGCM, the average diagnostic ability of the endoscopists was significantly improved in terms of accuracy, sensitivity, PPV, and NPV ($P < .05$).

Conclusions: EGCM exhibited comparable performance with senior endoscopists in the diagnosis of EGC and showed the potential value in aiding and improving the diagnosis of EGC by endoscopists. (Gastrointest Endosc 2021;93:1333-41.)

(footnotes appear on last page of article)

Gastric cancer (GC) is one of the most prevalent malignant carcinomas worldwide, with an estimated 1.0 million new cases per year. The prognosis of advanced GC is poor, whereas the 5-year survival rate of early GC (EGC) is more than 90% because of fewer lymph node metastases and a high curative endoscopic resection rate.[1] Therefore, a timely and accurate diagnosis of EGC is of great importance. However, the identification of EGC is quite challenging. The average detection rate of EGC reported in China is generally about 2% to 5%,[2] and the miss rate of EGC during gastroscopy is about 10%.[3] Considering the high incidence of GC, the missed number of EGC is overwhelming.

To overcome the current situation, magnifying endoscopy narrow-band imaging (ME-NBI) has been introduced by exploiting 2 discrete narrow bands of light (blue at 415 nm and green at 540 nm). ME-NBI can display the visualization of superficial mucosal and vascular structures, and several comparative studies have shown the advantages of ME-NBI over conventional white-light imaging on EGC detection.[4,5] It is currently the most powerful tool in evaluating EGC by analyzing the microstructure and microvessels. However, in real clinical practice, ME-NBI has not played its due role. Diagnostic performance of ME-NBI to differentiate EGC from noncancerous lesions requires substantial experience only found in expert centers. Therefore, the results in everyday clinical practice are commonly disappointing. Furthermore, diagnosis often varies widely between endoscopists, even within expert endoscopists.[6] The reported overall sensitivity of ME-NBI for the diagnosis of EGC ranged from 60% to 100% and the specificity

from 84% to 100%.[7-12] Therefore, it is appealing to develop an automatic predictive tool to assist in diagnosing EGC with high efficiency.

Today, artificial intelligence (AI) has great potential to aid decision-making in various medical fields, which could help nonexperts find abnormalities often missed.[13-19] Preliminary research has demonstrated the important values of AI on conventional white-light endoscopy regarding GC detection,[20,21] prediction of invasion depth,[22,23] and *Helicobacter pylori* infection.[24] However, the clinical applicability and reliability of AI based on ME-NBI remain questionable because of single-centered validation and lack of comparison with endoscopists.[25,26] Furthermore, the role of AI in improving ME-NBI recognition by endoscopists has not yet been investigated.

In this study, we developed and evaluated a computer-aided diagnosis model of EGC (EGCM) using ME-NBI images from 3 hospitals. We tested and compared the ability of EGCM and endoscopists to identify EGC based on ME-NBI and the performance of EGCM in delineating lesion boundaries and explored its role in aiding and improving the endoscopist's ME-NBI diagnosis of EGC.

METHODS

Data acquisition

Static endoscopic images of patients who underwent ME-NBI for the routine evaluation of a suspicious gastric lesion at the Endoscopic Center of Zhongshan Hospital (FDZS), The Affiliated Dongnan Hospital of Xiamen University (XMDN), and the Central Hospital of Wuhan (WH) between January 2017 and March 2020 were retrospectively obtained from the database. In the protocol of our hospitals, each patient is asked to sign an informed consent, which involves the donation of biologic samples and the use of health-related information for medical research before all endoscopies. Accordingly, patients undergoing the protocol provided their consent to using their deidentified images for public and nonprofit medical education and research. The study was approved by the Institutional Review Board of Fudan University (20180511). All ME-NBI images were taken at full magnification by standard endoscopes (GIF-H260Z or GIF-H290Z; Olympus Corp, Tokyo, Japan) with the NBI function, and the water immersion technique was adopted. ME-NBI images were included if the patient met these inclusion criteria: age $\geq$18 years and a definitive pathologic diagnosis. Images were excluded if the patient had multiple lesions, had no definitive pathologic diagnosis, and images were too blurry to be evaluated by senior endoscopists because of inadequate focus, halation, excessive bubbles, or bleeding.

ME-NBI images were annotated as EGC or noncancerous lesions by 3 experienced endoscopists (>10 years of experience) according to vessels plus surface classification and the MESDA-G diagnostic flow based on the pathologic results.[27] Initially, the annotation was made by 2 endoscopists. If there was disagreement between the 2 endoscopists regarding any images, a third endoscopist settled the dispute. The macroscopic type was classified based on the Paris classification[28,29] as elevated type (0-I, 0-IIa), flat/depressed type (0-IIb, 0-IIc, 0-III), and mixed type (a combination of elevated and flat/depressed lesion, 0-IIa+IIc, or other mixed types). Meanwhile, 2 highly experienced pathologists who were blinded to the ME findings performed a histologic evaluation according to the revised Vienna classification of GI epithelial neoplasia.[30] Category 3 (high-grade adenoma), 4 (intramucosal carcinoma), and 5 (submucosal invasion by carcinoma) were diagnosed as EGC, whereas category 1 (negative for neoplasia) and category 2 (mucosal low-grade neoplasia) were diagnosed as a noncancerous lesion. If there was disagreement, a reassessment was carried out to reach a consensus.

Model construction and validation

To identify EGC, a deep learning model was constructed within Pytorch (Facebook, Menlo Park, Calif, USA) based on the VGG-19 architecture. VGG-19 is a variant of the VGG model (Visual Geometry Group [VGG], Oxford University, Oxford, UK) that in short consists of 19 layers (16 convolution layers, 3 fully connected layers, 5 MaxPool layers, and 1 SoftMax layer).[31] The model was pretrained on the ILSVRC-2012 datasets (provided by the Large Scale Visual Recognition Challenge 2012, http://image-net.org/challenges/LSVRC/2012/) including 1.2 M training images, and a single fully connected 2-classification layer was designed to replace the top layer as the new top layer after the stack of convolutional layers.

FDZS patients were randomly assigned to the training cohort (TC) and internal test cohort (ITC) with a ratio of 7:3. Meanwhile, the TC was randomly divided into 2 subgroups with a ratio of 6:1, of which 6/7 of the (TCsub1) was used to train the model and the other 1/7 (TCsub2) was used to confirm the final parameters of the model. Before being entered into the network, the input images were normalized via a Z-score with the mean of (.569, .398, .319) and the standard deviation of (.154, .139, .120) that were automatedly computed from TCsub1 and were resized to 224 $\times$ 224 pixels to match the model. For each epoch, every 16 images were divided as a batch to be placed into the model. To optimize the model, the cross-entropy loss was used during the training process. Adam optimizer with an initial learning rate of 10^{-6} and a weight decay of .01 was used. To avoid overfitting, the validation was conducted on TCsub2 every 10 iterations, and an early stopping strategy was carried out with a patience of 10, which means the checkpoint would be selected as the final model when the loss of TCsub2 was no longer reduced for 10 consecutive times.

The robustness of the model was subsequently assessed using an external test cohort (ETC) from the other 2

hospitals (XMDN and WH). Also, the gradient-weighted class activation mapping (Grad-CAM) method was used in the ITC to visualize the area with the most outstanding contribution to the EGCM's prediction, which is helpful to show the appropriate trust in EGCM predictions.

Comparison between EGCM and endoscopists

The performance of the EGCM was compared with that of the endoscopists. Eight endoscopists with varying experience from 3 hospitals were divided into senior and junior groups according to their diagnostic experience. Specifically, 3 endoscopists (not involved in the annotations, each from 3 participating centers) with more than 10 years of ME-NBI experience were classified as the senior group, and 5 endoscopists with 12 to 24 months of NBI experience were classified as the junior group. Initially, both senior and junior endoscopists were asked to review the same ITC and make a diagnostic judgment (EGC or non-EGC) without knowing any clinical information or pathologic results. To explore the EGCM assistance ability, after at least 2 weeks all endoscopists were asked to make a diagnostic decision again on the same ITC with reference to the predicted probability given by the EGCM but not to the feedback of the first judging. The performances of the endoscopists with and without EGCM assistance were then compared.

Statistical analysis

The area under the receiver operating characteristic curve (AUC), accuracy, specificity, sensitivity, positive predictive value (PPV), and negative predictive value (NPV) were calculated to evaluate the performance of the model. The optimum cutoff point was determined in the TC using the Youden index. The McNemar test was applied to compare the difference in accuracy, sensitivity, and specificity among the different groups, whereas the χ^2 test was used to compare the PPV and NPV between different groups. Two-sided statistical tests were conducted, and statistical significance was determined when $P < .05$. Analyses were performed using R software for Windows (version 3.5.1; https://www.r-project.org) and Python (version 3.7.3; https://www.python.org).

RESULTS

Patient and clinical characteristics

Two hundred ninety-five patients (EGC, 128; noncancerous lesions, 167) and 1777 images from 3 different hospitals (FDZS, XMDN, and WH) around China were collected and analyzed. The patients from FDZS were randomly assigned to the TC and ITC with a 7:3 ratio. The robustness of the model was assessed using the ETC from the other 2 hospitals (XMDN and WH). Table 1 summarized the clinicopathologic characteristics of enrolled patients and lesions.

Performance of the EGCM

Based on the early stopping strategy, the final model was decided (iteration = 425, epoch = 7) when the loss in the validation cohort had not decreased for 10 consecutive times, as shown in Figure 1. The final model obtained an accuracy of .843, sensitivity of .870, and specificity of .819 in the TC. Its predictive performance was evaluated in the ITC and yielded a good evaluation result (accuracy, .770; sensitivity, .792; specificity, .745). The quantitative metrics are shown in Table 2.

Figure 2A shows the receiver operating characteristic curve analysis of the EGCM with an AUC of .808 in the ITC. The judgments of 8 endoscopists are also shown on the curve. On the ETC, the model also acquired a satisfactory performance (Fig. 2B; AUC, .813; accuracy, .763; sensitivity, .782; and specificity, .741).

Using Grad-CAM, we found that our model successfully distinguished and highlighted the abnormal areas. Figure 3 shows representative endoscopic images and corresponding activation mapping from the ITC.

Comparison between endoscopists and EGCM

The comparison results of the diagnostic performance between endoscopists and EGCM are shown in Table 3, and the detailed results of the endoscopists and the significance tests are shown in Supplementary Table 1 and Supplementary Table 2 (available online at www.giejournal.org), respectively. Both the receiver operating characteristic curve (Fig. 2A) and the quantitative metrics (Supplementary Fig. 1, available online at www.giejournal.org) showed the apparent group-based diagnostic discrepancies between senior and junior endoscopists. The results showed that the 3 senior endoscopists had better diagnostic abilities, with an average diagnostic accuracy of .755, whereas the junior endoscopists had a lower average diagnostic accuracy of .728. However, all subgroups and endoscopists had lower accuracy and sensitivity than the EGCM. It is worth noting that the EGCM obtained similar predictive performance as the senior endoscopists (accuracy: .770 vs .755, $P = .355$; sensitivity: .792 vs .767, $P = .183$; specificity: .745 vs .742, $P = .931$; PPV: .772 vs .764, $P = .515$; NPV: .767 vs .745, $P = .162$) (Supplementary Table 2), and their accuracy, sensitivity, and NPV were all significantly higher than their counterparts in the junior group and all endoscopists ($P < .05$)

Furthermore, we attempted to analyze sources of error in the diagnosis by the EGCM and endoscopists. To avoid too many confounding factors, we analyzed 2 special cases: where the model was incorrectly diagnosed and all physicians were correct, and where the model was correctly diagnosed and all physicians were incorrect. The statistical results (Table 4) showed that 6 of 7 in the first type were EGCs accounting for 85.7%, and the other case was a nonatrophic gastritis. After analyzing the characteristics of

TABLE 1. Clinical characteristics in the training and validation cohorts

Variable	Training cohort (n = 170) N (%)	Internal test cohort (n = 73) N (%)	External test cohort (n = 52) N (%)
Age, y	59 (53-66)	62 (50-66)	55 (49-62)
Gender			
Male	106 (62.35)	46 (63.01)	28 (53.85)
Female	64 (37.65)	27 (36.99)	24 (46.15)
Tumor location			
Upper 1/3	13 (7.65)	12 (16.44)	7 (13.46)
Middle 1/3	51 (30)	20 (27.40)	10 (19.23)
Lower 1/3	106 (62.35)	41 (56.16)	35 (67.31)
Morphology			
Elevated type (0-I,0-IIa)	54 (31.76)	18 (24.66)	10 (19.23)
Flat/depressed type (0-IIb,0-IIc,0-III)	86 (50.59)	41 (56.16)	24 (46.15)
Mixed type (0-IIa+IIc,etc)	30 (17.65)	14 (19.18)	18 (34.62)
Pathology			
Negative for neoplasia	79 (46.47)	30 (14.10)	29 (55.77)
Mucosal low-grade neoplasia	19 (11.18)	5 (6.85)	5 (9.62)
High-grade adenoma or intramucosal carcinoma	59 (34.71)	30 (41.10)	17 (32.96)
Submucosal invasion by carcinoma	13 (7.65)	8 (10.96)	1 (1.92)

Values are median (interquartile range) or n (%).

these ME-NBI pictures of EGC, we found that the common feature of these ME-NBI pictures was the small percentage of cancerous areas, small focal carcinoma, or cancerous tissue intersecting with normal tissue (Supplementary Fig. 2, available online at www.giejournal.org). For the second type, 24 were correctly diagnosed by the EGCM but were incorrectly diagnosed by the endoscopists (Supplementary Fig. 3, available online at www.giejournal.org). Nonatrophic gastritis was the predominant type (21/24, 87.5%), and analysis of these ME-NBI images revealed that these lesions mainly appeared as erosive gastritis with whitish-grayish patches, little hemorrhage, or surrounding ductal dilatation.

As shown in Table 3 and Figure 4A, the endoscopists achieved significantly improved average performance after referring to EGCM results ($P < .05$). Subgroup analysis based on the different groups of endoscopists (Table 3, Fig. 4B) further showed that with EGCM assistance, the average accuracy, specificity, and PPV of the junior endoscopists were significantly improved to .747 ($P < .05$), .813 ($P < .05$), and .800 ($P < .05$) respectively, which is comparable with the diagnostic performance of the senior endoscopists without the assistance of the EGCM. On the other hand, the performance of the senior group was also significantly improved with the assistance of the EGCM (accuracy: .755 vs .789, $P < .05$; sensitivity: .767 vs .874, $P < .05$; NPV: .745 vs .836, $P < .05$) with significantly better performance than EGCM in accuracy, sensitivity, specificity, and NPV ($P < .05$).

DISCUSSION

In this present study, we used a deep learning VGG-19 architecture to develop an AI-based EGC diagnostic system, named EGCM, which was trained and validated using 1777 ME-NBI images from 3 different hospitals. To the best of our knowledge, this is the first study to evaluate AI in spotting suspected EGC based on ME-NBI images using a multicenter validation cohort. It is also the first to demonstrate that AI could improve the endoscopists' diagnosis of EGC with ME-NBI images.

In our study, all morphologic types of EGC in the Paris classification were covered in all cohorts, including type 0-I, type 0-III, and mixed type. Meanwhile, we kept some of the bleeding and mildly blurred images. This guaranteed that the included images for training and testing EGCM tend to be closer to the "real world." Our EGCM showed satisfactory performance in both ITC and ETC, which showed the generalized ability of our model to other centers with different distributions of the data.

We compared the ability of EGCM and endoscopists to identify EGC, and the comparison results showed that EGCM had a similar capability to that of senior endoscopists in terms of accuracy, sensitivity, specificity, PPV, and NPV ($P > .05$). In addition, the EGCM model showed significantly better performance than the junior endoscopists ($P < .05$). Most importantly, EGCM could effectively improve the diagnostic performance of endoscopists of

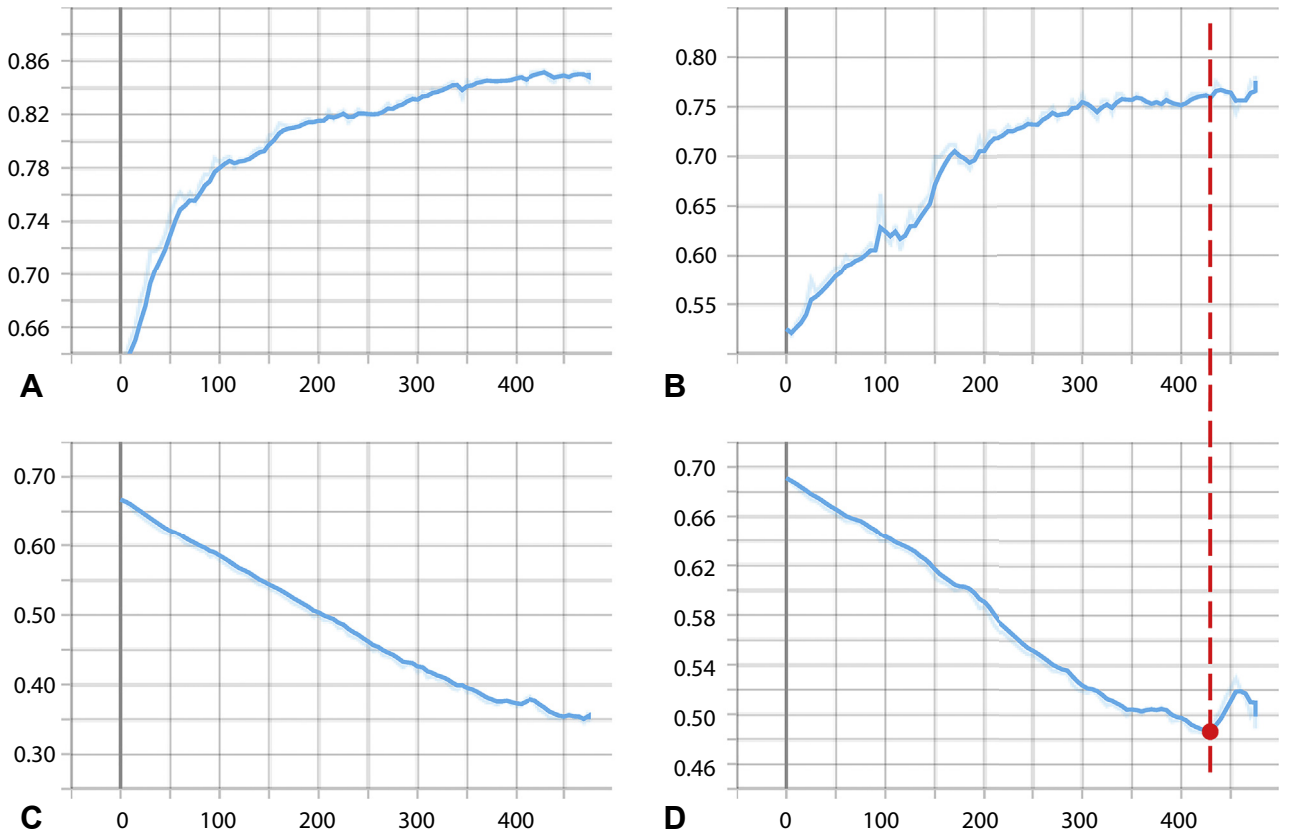


Figure 1. Accuracy and loss curves (smoothing factor = .6) during the training process using TensorBoard (Software available from tensorflow.org, Google, Mountain View, Calif, USA) where the *red point* means the early stopping point. **A,** Accuracy curve of TCsub1 (6/7 of the training cohort). **B,** Accuracy curve of TCsub2 (1/7 of the training cohort). **C,** Loss curve for TCsub1. **D,** Loss curve for TCsub2.

TABLE 2. Performance of the computer-aided early gastric cancer diagnosis model

Cohort	Area under the receiver operating characteristic curve (95% confidence interval)	Accuracy	Sensitivity	Specificity	Positive predictive value	Negative predictive value
Endoscopic Center of Zhongshan Hospital						
Training cohort	.914 (.897-.930)	.843	.870	.819	.806	.879
Internal test cohort	.808 (.769-.846)	.770	.792	.745	.772	.767
Central Hospital of Wuhan and Affiliated Dongnan Hospital of Xiamen University						
External test cohort	.813 (.751-.876)	.763	.782	.741	.782	.741

different levels. After referring to the diagnosis results of the EGCM, the diagnostic accuracy of the junior endoscopists was significantly improved from .728 to .747 ($P < .05$) and became comparable with the senior endoscopists. Similarly, the performance of senior endoscopists was also significantly improved and was significantly better than EGCM alone in terms of accuracy, sensitivity, and specificity ($P < .05$). These results demonstrate the additional value of the EGCM in the auxiliary diagnosis of EGC. With cues suggested by the EGCM, inexperienced endoscopists can continuously improve the accuracy of disease recognition and gain experience for correct identification of lesions. It is of great importance for less-experienced users of ME in poor areas.

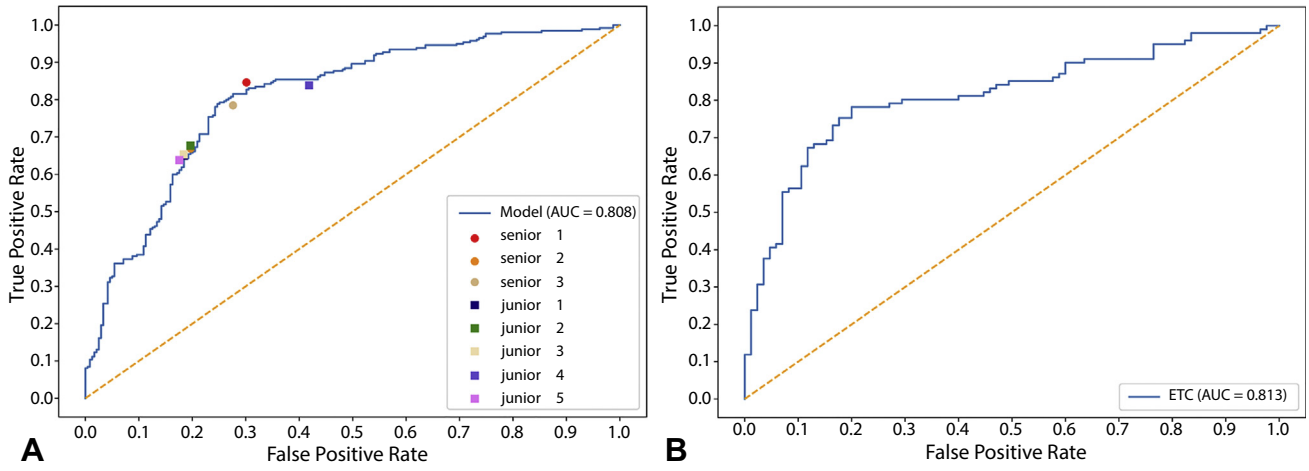


Figure 2. The area under the receiver operating characteristic curves (AUCs) of the computer-aided early gastric cancer diagnosis model (EGCM) in the internal test cohort (ITC) and external test cohort (ETC). **A,** AUC analysis of the EGCM and endoscopists in the ITC. **B,** AUC analysis of the EGCM in the ETC.

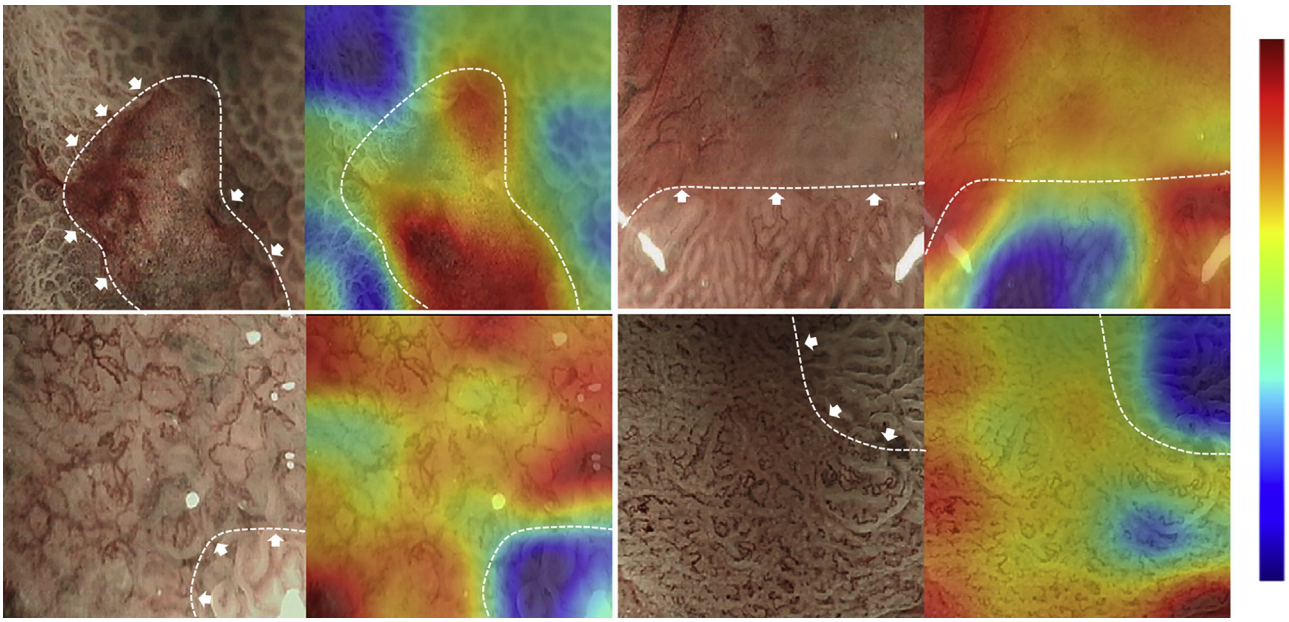


Figure 3. Examples of class activation maps for the model using gradient-weighted class activation mapping. The areas shown in *red* are abnormal areas, indicated by *white arrows* and *dashed white lines*.

Another advantage of the EGCM is its ability to delineate lesion boundaries based on ME-NBI images. On the one hand, it can assist the endoscopist in focusing on suspicious areas and improve the detection rate of EGC. On the other hand, it can help identify the endoscopist's biopsy sampling site, improve the positive rate of biopsy sampling, and, more importantly, help to spot more serious lesions. This is especially important to avoid and reduce endoscopic treatment on inappropriate lesions. At present, only the support vector machine model reported by Kanesaka et al[26] was able to delineate the scope of lesions based on ME-NBI images. Compared with the former study requiring manual annotation of the extent of lesions for ME-NBI images, our model is based on unsupervised learning of neural networks, and depending on the Grad-CAM scoring, it automatically annotates highly suspicious areas on images. Based on the characteristics of the convolutional neural network design, the accuracy of EGCM delineation may further increase with the increase in learning data.

In the analysis of misdiagnosis by EGCM and endoscopists, we found that errors in the EGCM diagnosis were mainly on EGC with a small area or with an intermingled growth pattern with normal tissues (Supplementary

TABLE 3. Diagnostic comparison between endoscopists and the EGCM

	Participant	Accuracy	Sensitivity	Specificity	Positive predictive value	Negative predictive value
Without EGCM	EGCM	.770	.792	.745	.772	.767
	Junior	.728*	.692*	.768	.764	.696*
	Senior	.755	.767	.742	.764	.745
	All endoscopists	.738*	.720*	.758	.764	.713*
With EGCM	Junior	.747*,†	.686*	.813*,†	.800*,†	.704*
	Senior	.789*,†	.874*,†	.696*,†	.758	.836*,†
	All endoscopists	.763*,†	.757*,†	.769*	.781†	.744*,†

EGCM, Computer-aided early gastric cancer diagnosis model.
*Significant difference between the target group and EGCM.
†Significant difference between the results of groups without EGCM and with EGCM.

TABLE 4. Distribution of cases with conflicting diagnoses*

	EGCM×, all endoscopists √	EGCM√, all endoscopists ×
Early gastric cancer	6	1
Nonatrophic gastritis	1	20
Atrophic gastritis	0	3
Total	7	24

EGCM, Computer-aided early gastric cancer diagnosis model; X, Incorrect prediction; √, correct prediction.
*Gastritis was classified according to the consensus on chronic gastritis in China[34] and the Updated Sydney System.[35]

Fig. 2), whereas misdiagnosis of endoscopists was mainly concentrated on erosive lesions (Supplementary Fig. 3). These results are very consistent with actual clinical practices and also reflect the different advantages of manual and intelligent diagnosis.[32] For a few cases of small cancerous areas or with an intermingled growth with normal tissues, the EGCM had a relatively mediocre performance. That may be because too much invalid information makes it difficult to extract enough representative features of imaging supporting EGCM decisions. But for erosive gastritis, the EGCM can correctly distinguish such cases, whereas endoscopists are prone to misdiagnosing them as early cancer. Considering that most common EGC is usually manifested by erosion and erosive gastritis is a common type of gastritis, it is not rare for the endoscopist to misdiagnose the erosion as EGC.[2,32] In this study, 23 cases of gastritis were misdiagnosed as EGC. These results are again evidence of the complementary roles of the endoscopist and EGCM diagnosis. Most importantly, the EGCM not only helps endoscopists improve diagnostic accuracy but also has significance in reducing endoscopic overtreatment and preventing patients from undergoing unnecessary ESD procedures.

Recent studies also attempted to develop an AI method in the diagnosis of ME-NBI of EGC. Kanesaka et al[26] developed a support vector machine software to differentiate EGC from noncancerous lesions, showing 97% sensitivity and 95% specificity. It heavily relied on manual design and delineation, which was time-consuming and difficult to generalize. Li et al[25] developed a convolutional neural network to analyze gastric mucosal lesions observed by ME-NBI and achieved an accuracy of 90.91%; however, the network either lacked comparative studies with endoscopists or had incomplete coverage of EGC morphologic type, in which type 0-I and type 0-III lesions were excluded. Luo et al[33] recently developed a deep learning semantic segmentation model to detect upper GI cancers from suspicious lesions. The differences are that their studies were based on high-resolution standard white-light images and manual labeling of the TC was involved, whereas our EGCM is based on ME-NBI for end-to-end unsupervised training.

This study has several limitations. First, the training data were collected from one center (FDZS). Training using images from multiple centers will help generalize the model and reduce performance fluctuations caused by different validation sets in a future study and for practical use. Second, there is a possible bias with the previous exposure of endoscopists to the same images, although participating endoscopists were not informed of the accuracy of each image in their first diagnosis. It is better to use completely different ME-NBI images for the performance improving study. Third, our EGCM was not satisfactory for rare types of EGCs, such as those with an intermingled growth pattern. Increasing the number of ME-NBI images of such rare EGCs would help improve the accuracy of the EGCM. Fourth, only static images were used for EGCM training in our study. Considering that the ultimate application scenario of AI is a real-time diagnosis based on video streams, a study using video data directly for EGC training is highly desirable. Fifth, our EGCM does not detect lesions with nonmagnified views. This means the endoscopist must suspect that a lesion exists first before being able to use the EGCM. The integration of EGCM with other developed computer-assisted devices on nonmagnified endoscopy could be a potential way around this issue.

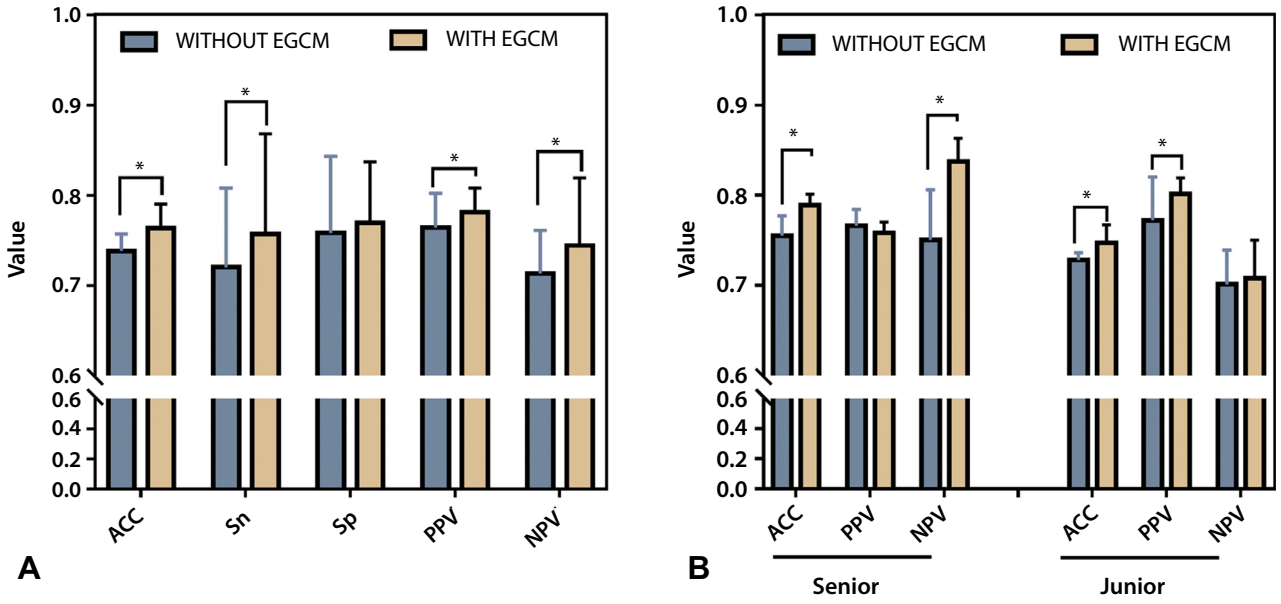


Figure 4. Comparison between endoscopists and the computer-aided early gastric cancer diagnosis model (EGCM). **A,** The average performance of endoscopists before and after EGCM assistance. **B,** Subgroup analysis of the performance differences before and after EGCM assistance. *ACC*, Accuracy; *Sn*, sensitivity; *Sp*, specificity; *PPV*, positive predictive value; *NPV*, negative predictive value.

Sixth, the training, testing, and validation were all done with the Olympus Z series. Hence, the results might not be reproducible for the digital/automatic zoom series and for other brands of endo scopes. Last, the EGCM only differentiates cancerous from noncancerous lesions and not the depth of invasion. Considering the depth is closely related to the choice of treatment, further development of a computer-assisted device that is capable of providing predictions of invasion depth is expected.

In summary, we constructed an EGCM model for assisting EGC diagnosis. The model exhibited comparable diagnostic capability with that of expert endoscopists and showed potential in improving the diagnostic performance of nonexpert endoscopists.

REFERENCES

1. Sano T, Katai H, Sasako M, et al. The management of early gastric cancer. Surg Oncol 2000;9:17-22.
2. Ren W, Yu J, Zhang Z, et al. Missed diagnosis of early gastric cancer or high-grade intraepithelial neoplasia. World J Gastroenterol 2013;19:2092-6.
3. Pimenta-Melo AR, Monteiro-Soares M, Libânio D, et al. Missing rate for gastric cancer during upper gastrointestinal endoscopy: a systematic review and meta-analysis. Eur J Gastroenterol Hepatol 2016;28:1041-9.
4. Yao K. Clinical application of magnifying endoscopy with narrow-band imaging in the stomach. Clin Endosc 2015;48:481-90.
5. Zhang Q, Wang F, Chen ZY, et al. Comparison of the diagnostic efficacy of white light endoscopy and magnifying endoscopy with narrow band imaging for early gastric cancer: a meta-analysis. Gastric Cancer 2016;19:543-52.
6. East JE, Vleugels JL, Roelandt P, et al. Advanced endoscopic imaging: European Society of Gastrointestinal Endoscopy (ESGE) technology review. Endoscopy 2016;48:1029-45.
7. Ezoe Y, Muto M, Horimatsu T, et al. Magnifying narrow-band imaging versus magnifying white-light imaging for the differential diagnosis of gastric small depressive lesions: a prospective study. Gastrointest Endosc 2010;71:477-84.
8. Ezoe Y, Muto M, Uedo N, et al. Magnifying narrowband imaging is more accurate than conventional white-light imaging in diagnosis of gastric mucosal cancer. Gastroenterology 2011;141:2017-25.
9. Maki S, Yao K, Nagahama T, et al. Magnifying endoscopy with narrow-band imaging is useful in the differential diagnosis between low-grade adenoma and early cancer of superficial elevated gastric lesions. Gastric Cancer 2013;16:140-6.
10. Nonaka K, Arai S, Ban S, et al. Prospective study of the evaluation of the usefulness of tumor typing by narrow band imaging for the differential diagnosis of gastric adenoma and well-differentiated adenocarcinoma. Dig Endosc 2011;23:146-52.
11. Tao G, Xing-Hua L, Ai-Ming Y, et al. Enhanced magnifying endoscopy for differential diagnosis of superficial gastric lesions identified with white-light endoscopy. Gastric Cancer 2014;17:122-9.
12. Yao K, Doyama H, Gotoda T, et al. Diagnostic performance and limitations of magnifying narrow-band imaging in screening endoscopy of early gastric cancer: a prospective multicenter feasibility study. Gastric Cancer 2014;17:669-79.
13. Dong D, Tang L, Li Z, et al. Development and validation of an individualized nomogram to identify occult peritoneal metastasis in patients with advanced gastric cancer. Ann Oncol 2019;30:431-8.
14. Bi W, Hosny A, Schabath M, et al. Artificial intelligence in cancer imaging: clinical challenges and applications. CA Cancer J Clin 2019;69:127-57.
15. Dong D, Fang M, Tang L, et al. Deep learning radiomic nomogram can predict the number of lymph node metastasis in locally advanced gastric cancer: an international multicenter study. Ann Oncol 2020;31:912-20.
16. Li W, Zhang L, Tian C, et al. Prognostic value of computed tomography radiomics features in patients with gastric cancer following curative resection. Eur Radiol 2019;29:3079-89.
17. Li J, Fang M, Wang R, et al. Diagnostic accuracy of dual-energy CT-based nomograms to predict lymph node metastasis in gastric cancer. Eur Radiol 2018;28:5241-9.
18. Zhang W, Fang M, Dong D, et al. Development and validation of a CT-based radiomic nomogram for preoperative prediction of early recurrence in advanced gastric cancer. Radiother Oncol 2020;145:13-20.

19. Zhang L, Dong D, Zhang W, et al. A deep learning risk prediction model for overall survival in patients with gastric cancer: a multicenter study. Radiother Oncol 2020;150:73-80.
20. Ishioka M, Hirasawa T, Tada T. Detecting gastric cancer from video images using convolutional neural networks. Dig Endosc 2019;31:e34-5.
21. Lee J, Kim Y, Kim Y, et al. Spotting malignancies from gastric endoscopic images using deep learning. Surg Endosc 2019;33:3790-7.
22. Kubota K, Kuroda J, Yoshida M, et al. Medical image analysis: computer-aided diagnosis of gastric cancer invasion on endoscopic images. Surg Endosc 2012;26:1485-9.
23. Zhu Y, Wang Q, Xu M, 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.
24. Itoh T, Kawahira H, Nakashima H, et al. Deep learning analyzes *Helicobacter pylori* infection by upper gastrointestinal endoscopy images. Endosc Int Open 2018;6:E139-44.
25. Li L, Chen Y, Shen Z, et al. Convolutional neural network for the diagnosis of early gastric cancer based on magnifying narrow band imaging. Gastric Cancer 2020;23:126-32.
26. Kanesaka T, Lee TC, Uedo N, et al. Computer-aided diagnosis for identifying and delineating early gastric cancers in magnifying narrow-band imaging. Gastrointest Endosc 2018;87:1339-44.
27. Muto M, Yao K, Kaise M, et al. Magnifying endoscopy simple diagnostic algorithm for early gastric cancer (MESDA-G). Dig Endosc 2016;28:379-93.
28. Participants in the Paris Workshop. The Paris endoscopic classification of superficial neoplastic lesions: esophagus, stomach, and colon: November 30 to December 1, 2002. Gastrointest Endosc 2003;58:S3-43.
29. Endoscopic Classification Review Group. Update on the Paris classification of superficial neoplastic lesions in the digestive tract. Endoscopy 2005;37:570-8.
30. Dixon MF. Gastrointestinal epithelial neoplasia: Vienna revisited. Gut 2002;51:130-1.
31. Simonyan K, Zisserman A. Very deep convolutional networks for large scale image recognition. arXiv 2015:1409-556.
32. Zhu L, Qin J, Wang J, et al. Early gastric cancer: current advances of endoscopic diagnosis and treatment. Gastroenterol Res Pract 2016;2016:9638041.
33. 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.
34. Fang J, Du Y, Liu W, et al. Chinese consensus on chronic gastritis (2017, Shanghai). J Dig Dis 2018;19:182-203.
35. Dixon M, Genta R, Yardley J, et al. Classification and grading of gastritis. The updated Sydney System. International Workshop on the Histopathology of Gastritis, Houston 1994. Am J Surg Pathol 1996;20:1161-81.

Abbreviations: AI, artificial intelligence; AUC, area under the receiver operating characteristic curve; EGC, early gastric cancer; EGCM, computer-aided early gastric cancer diagnosis model; ETC, external test cohort; FDZS, Endoscopic Center of Zhongshan Hospital; GC, gastric cancer; Grad-CAM, gradient-weighted class activation mapping; ITC, internal test cohort; ME-NBI, magnifying endoscopy narrow-band imaging; NPV, negative predictive value; PPV, positive predictive value; TC, training cohort; WH, Central Hospital of Wuhan; XMDN, Affiliated Dongnan Hospital of Xiamen University.

DISCLOSURE: *All authors disclosed no financial relationships. Research support for this study was provided by the Natural Science Foundation of Shanghai (grant no. 18411952500, Hao Hu and Pinghong Zhou) and Shanghai Municipal Human Resources Development Program for Outstanding Young Talents in Medical and Health Sciences (grant no. 2018YQ33, Hao Hu), National Natural Science Foundation of China (grants nos. 91959130 [Di Dong], 81971776 [Di Dong], 81771924 [Di Dong], 81930053 [Jie Tian], 81701750 [Hao Hu], and 81900548 [Min Wang]), Smart Medical Program of Shanghai Municipal Health Commission (grant no. 2018ZHYL0204, Pinghong Zhou and Hao Hu), the National Key R&D Program of China (grant nos. 2017YFA0700401 [Xin Yang], 2017YFC1309100 [Di Dong], 2017YFA0205200 [Jie Tian], and 2017YFC1308700 [Di Dong]), the Beijing Natural Science Foundation (grant no. L182061, Di Dong), Strategic Priority Research Program of Chinese Academy of Sciences (XDB 38040200, Di Dong), and the Youth Innovation Promotion Association CAS (grant no. 2017175, Di Dong).*

*Drs Hu and Dong and Ms Gong contributed equally to this article.

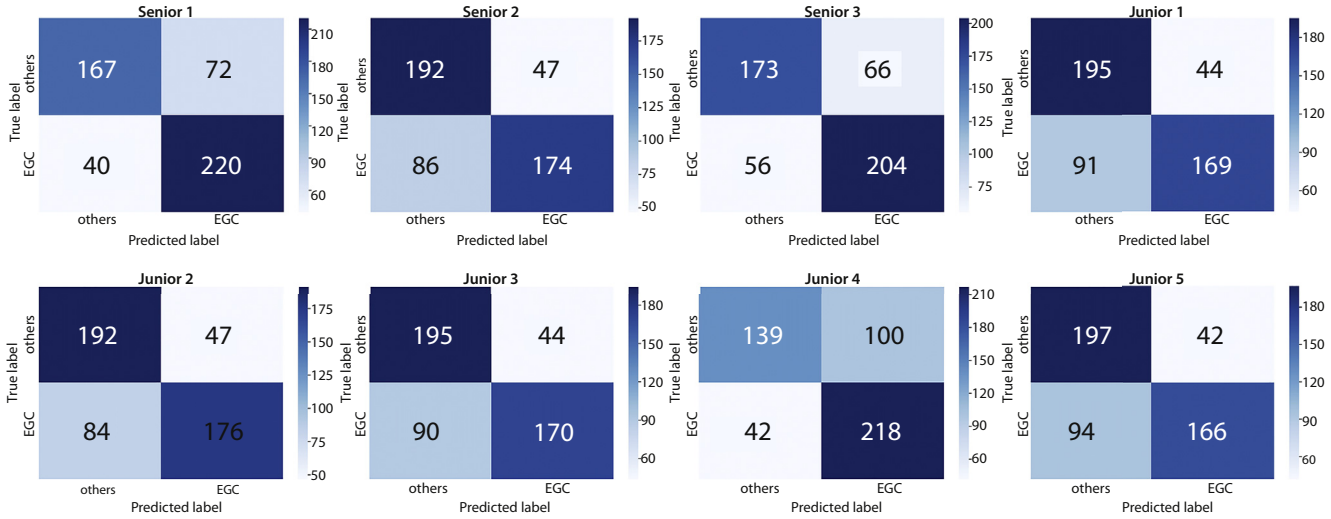
0016-5107/$36.00
https://doi.org/10.1016/j.gie.2020.11.014

Received September 14, 2020. Accepted November 18, 2020.

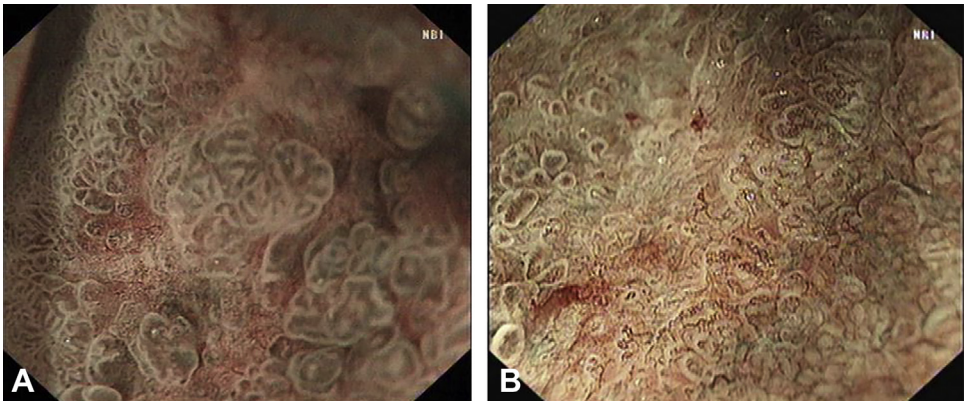
Current affiliations: Endoscopy Center and Endoscopy Research Institute, Zhongshan Hospital, Fudan University, Shanghai, China (1), College of Medicine and Biological Information Engineering School, Northeastern University, Shenyang, China (2), CAS Key Laboratory of Molecular Imaging, Beijing Key Laboratory of Molecular Imaging, The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China (3), School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China (4), Department of Gastroenterology, Hepatology and Nutrition, Shanghai Children's Hospital, Shanghai Jiaotong University, Shanghai, China (5), Endoscopy Center, Zhongshan Hospital (Xiamen Branch), Fudan University, Xiamen, China (6), Department of Gastroenterology, No. 1 Hospital of Wuhan, Wuhan, China (7), Department of Gastroenterology, The Affiliated Dongnan Hospital of Xiamen University, Zhangzhou, China (8), Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, School of Medicine and Engineering, Beihang University, Beijing, China (9).

Reprint requests: Jie Tian, PhD, Director of the CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, No. 95, Zhongguancun East Rd, Beijing, 100190, China, and Pinghong Zhou, MD, Endoscopy Center and Endoscopy Research Institute, Zhongshan Hospital of Fudan University, 180 Fenglin Road, Xuhui District, Shanghai, 200032, China.

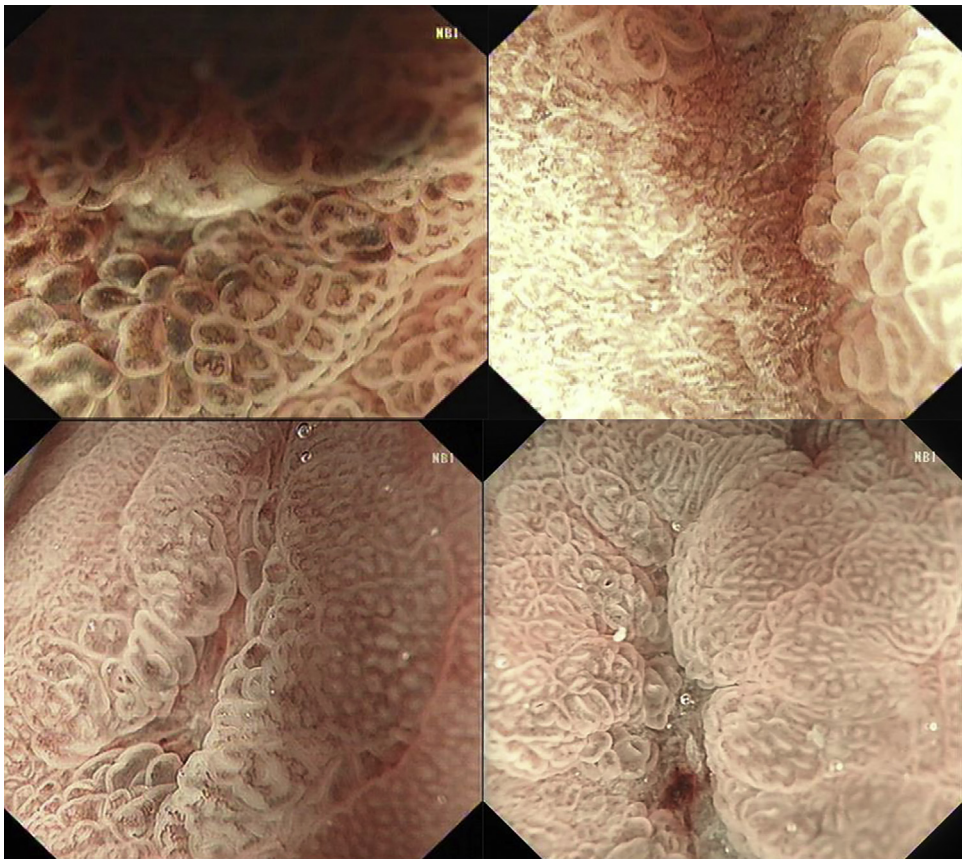
If you would like to chat with an author of this article, you may contact Dr Tian at tian@ieee.org or Dr Zhou at zhou.pinghong@zs-hospital.sh.cn.



Supplementary Figure 1. Diagnostic results of 8 independent endoscopists in the internal test cohort. *EGC*, Early gastric cancer.



Supplementary Figure 2. Representative magnifying endoscopy narrow-band images of early gastric cancer. The cancerous tissue grows intermingled with normal tissue. **A,** A 46-year-old man with undifferentiated adenocarcinoma, carcinoma located in the lamina propria of the mucosa, and focal invasion into the mucosal muscular layer. **B,** A 65-year-old man with adenocarcinoma with grade II differentiation and carcinoma tissue invading the mucosal muscle layer. *NBI*, Narrow-band imaging.



Supplementary Figure 3. Representative magnifying endoscopy narrow-band images of erosive gastritis with whitish-grayish patches, little hemorrhage, or ductal dilatation. *NBI*, Narrow-band imaging.

SUPPLEMENTARY TABLE 1. Performance of endoscopists

Endoscopist no.	Without EGCM					With EGCM				
	Accuracy	Sensitivity	Specificity	PPV	NPV	Accuracy	Sensitivity	Specificity	PPV	NPV
Senior										
Senior 1	.776	.846	.699	.753	.807	.792	.900	.674	.750	.861
Senior 2	.733	.669	.803	.787	.691	.776	.850	.695	.752	.810
Senior 3	.756	.785	.724	.756	.755	.800	.873	.720	.772	.839
Junior										
Junior 4	.729	.650	.816	.793	.682	.733	.650	.824	.801	.684
Junior 5	.737	.677	.803	.789	.696	.749	.658	.849	.826	.695
Junior 6	.731	.654	.816	.794	.684	.733	.658	.816	.795	.687
Junior 7	.715	.838	.582	.686	.768	.780	.808	.749	.778	.782
Junior 8	.727	.638	.824	.798	.677	.739	.658	.828	.807	.690

EGCM, Computer-aided early gastric cancer diagnosis model; *PPV*, positive predictive value; *NPV*, negative predictive value.

SUPPLEMENTARY TABLE 2. Results of a significant test

	Accuracy	Sensitivity	Specificity	Positive predictive value	Negative predictive value
Without EGCM					
Junior vs EGCM	*	*	.080	.487	*
Senior vs EGCM	.355	.183	.931	.515	.162
All vs EGCM	*	*	.195	.343	*
With EGCM					
Junior vs EGCM	*	*	*	*	*
Senior vs EGCM	*	*	*	.151	*
All vs EGCM	*	*	*	.141	*
EGCM's guiding ability					
Junior	*	.616	*	*	.205
Senior	*	*	*	.570	*
All endoscopists	*	*	.179	*	*

EGCM, Computer-aided early gastric cancer diagnosis model; *All vs EGCM*, significant test results between all endoscopists and the EGCM.
*$P < .05$.