Check for updates

A system based on deep convolutional neural network improves the detection of early gastric cancer

Jie Feng[1,2], Shang rui Yu[1], Yao ping Zhang[1,2], Lina Qu[1,2], Lina Wei[1], Peng fei Wang[1], Li juan Zhu[3], Yanfeng Bao[3], Xiao gang Lei[4], Liang liang Gao[5], Yan hu Feng[1], Yi Yu[1] and Xiao jun Huang[1,2]*

[1]Department of Gastroenterology, Lanzhou University Second Hospital, Lanzhou, Gansu, China, [2]Technology Research and Development Department, Digestive Endoscopy Engineering Research and Development Center of Gansu Province, Lanzhou, Gansu, China, [3]Department of Sciences and Technology, Beijing Huag gen Anbang Technology Technology Company Limited, Beijing, China, [4]Department of Gastroenterology, Lanzhou Cheng guan District People's Hospital, Lanzhou, Gansu, China, [5]Department of Gastroenterology, Min County People's Hospital, Ding Xi, Gansu, China

Background: Early gastric cancer (EGC) has a high survival rate, but it is difficult to diagnosis. Recently, artificial intelligence (AI) based on deep convolutional neural network (DCNN) has made significant progress in the field of gastroenterology. The purpose of this study was to establish a DCNN assist system to improve the detection of EGC.

Methods: 3400 EGC and 8600 benign images were collected to train the DCNN to detect EGC. Subsequently, its diagnostic ability was compared to that of endoscopists using an independent internal test set (ITS, including 1289 images) and an external test set (ETS, including 542 images) come from three digestive center.

Results: The diagnostic time of DCNN and endoscopists were 0.028s, 8.05 $\pm$ 0.21s, 7.69 $\pm$ 0.25s in ITS, and 0.028s, 7.98 $\pm$ 0.19s, 7.50 $\pm$ 0.23s in ETS, respectively. In ITS, the diagnostic sensitivity and accuracy of DCNN are 88.08%(95% confidence interval,95%CI,85.24%-90.44%), 88.60% (95% CI,86.74%-90.22%), respectively. In ETS, the diagnostic sensitivity and accuracy are 92.08% (95%CI, 87.91%- 94.94%),92.07%(95%CI, 89.46%-94.08%), respectively. DCNN outperformed all endoscopists in ETS, and had a significantly higher sensitivity than the junior endoscopists(JE)(by18.54% (95% CI, 15.64%-21.84%) in ITS, also higher than JE (by21.67%,95%CI, 16.90%-27.32%) and senior endoscopists (SE) (by2.08%, 95%CI, 0.75%-4.92%)in ETS. The accuracy of DCNN model was higher (by10.47%,95%CI, 8.91%-12.27%) than that of JE in ITS, and also higher (by14.58%,95%CI, 11.84%-17.81%; by 1.94%,95%CI,1.25%-2.96%, respectively) than JE and SE in ETS.

Conclusion: The DCNN can detected more EGC images in a shorter time than the endoscopists. It will become an effective tool to assist in the detection of EGC in the near future.

Introduction

According to the 2020 Global Cancer Statistics, Gastric cancer is the third most lethal and the fifth most common malignancy from a global perspective, and Asia remains the region with a high incidence of cancer, with a cancer incidence rate of 49.3% and a mortality rate of 58.3%, with 719524 new cases of gastric cancer (1). The survival rate of patients with stage IA was 91%, Whereas the patients with stage IV less than 17% (2) Therefore, the early detection of gastric cancer is particularly important. However, the diagnosis of early gastric cancer (EGC) is difficult and often be ignored, especially in countries with large populations, such as China: the detection rate of EGC in China is only 10%, much lower than that in South Korea (50%) and Japan (70%) (3, 4), the diagnosis rate of EGC has great room for improvement. However, the large number of patients, insufficient diagnostic knowledge and experience of physicians, lack of advanced endoscopic equipment, and shortage of endoscopists have seriously affected the improvement of the diagnostic level of EGC in China. These problems are particularly prominent in primary medical institutions (5).

Some studies have reported a false negative rate of 4.6-25.8% in the detection of gastric cancer by esophagogastroduodenoscopy (EGD) (6–14), 71.4% of gastric cancer patients were initially diagnosed with gastritis, ulcers or "suspicious lesions", with the majority (73%) of errors made by endoscopists (9), technical factors and subjective cognition have significant influence on the screening of EGC (15). The detection of EGC requires not only well-trained endoscopists but also comprehensive knowledge (16), secondly, it is also necessary for endoscopists to avoid the influence of subjective factors, which limit the detection of EGC (17). Therefore, it is very important to develop a tool that has good detection ability and will not be affected by subjective factors to assist endoscopists in the detection of EGC. In recent years, artificial intelligence (AI) based on deep convolutional neural deep learning (DCNN) has come into being, and DCNN has made remarkable progress in various fields, including medicine. In the field of digestive endoscopy, it has been applied to the detection of colonic polyps (18) and the diagnosis of auxiliary capsule endoscopy (19). Based on the above reasons, we constructed an auxiliary diagnosis system for EGC base on DCNN, and tested the diagnostic efficiency of DCNN, aiming to improve the diagnostic efficiency of EGC.

Methods

Training dataset preparation

The DCNN was trained using EGD images obtained from the digestive center of Lanzhou University Second Hospital, total 12000 images were selected from the database from January, 2013 to December 2019, including 3400 images of EGC, 8600 images of benign lesions and normal images. All the lesions included in the study were confirmed by biopsy or surgical pathology and the lesion scope was clear, the patient and lesion characteristics of EGC in training set was shown in Supplementary Material 1. Postoperative pathological diagnosis included high-grade intraepithelial neoplasia and carcinoma confined to mucosa or submucosa. The equipment for endoscopic images was standard GIF-GIF-260/H290Z, Olympus Medical Systems, Co., Ltd., Tokyo, Japan) and a standard endoscopic video system (EVIS LUCERA ELITE CV-290/CLV-290SL; Olympus Medical Systems), and all images were white light endoscopes without magnification. Images containing poor inflation, halo, blur, defocus or mucus, and post-biopsy bleeding were excluded from the training dataset.

DCNN training

According to the outcome of pathology, the EGD images were labeled as EGC and other benign lesions, computer engineers will be annotated images for unified clipping, color space transformation, denoising, image morphology operations and normalization of a series of processing, eliminate human and environmental interference, better display image features, enhance the robustness of the algorithm. Algorithm engineers used the DCNN module to test multiple computer models such as DLA34 and Swim Transformer Tiny, and put the training set into the model for training. Through observation and comparison, the 18-layer convolutional neural network model

with the optimal accuracy and speed was determined. Input resolution: 512 × 512, batch-size: 32, initial learning rate: 1.25e-4, optimization: Adam

DCNN testing

We used standard EGC images independent of the training set to verify the accuracy of DCNN (From January 2020 to October 2020). Our test set was divided into internal test set (ITS) and external test set (ETS). The images of the ETS were from Wuwei Cancer Hospital and Minxian People's Hospital. The test set excluded postoperative gastric images, magnification, staining endoscopy, mucus and halo images. The patient and lesion characteristics of EGC in ITS and ETS set was shown in Table 1. The ITS contains 1289 images (604 EGC) and the ETS contains 542 images (240 EGC). Non-cancer images include ulcers, polyps and chronic gastritis. We identified each lesion area by comparing endoscopic images with the extent of the lesion in the excised specimen, and manually annotated all gastric cancer lesions in the test data set by two experienced endoscopists (L.W and P.W) using a true red rectangular border.

TABLE 1   Patient and lesion characteristics of EGC in Internal test set and External test set.

Patient characteristics	Internal test set (n=134)	External test set (n=48)
Age (years), Mean (range)	60 (27–79)	65 (35-84)
Sex, Number (%)		
Male	92 (68.66)	38 (79.17)
Female	42 (31.34)	10 (20.83)
Size of lesion (mm), Median (range)	15 (4-48.5)	17.4 (6-40.2)
Tumor location, Number (%)		
Cardia$_s$ Fundus of the stomach,	18 (13.43)	3 (6.25)
Number (%)		
Body	58 (43.28)	13 (27.08)
Angle	28 (20.90)	12 (25)
Antrum	30 (22.39)	20 (41.67)
Macroscopic type, Number (%)		
0-1	8 (5.97)	1 (2.08)
O-IIa	33 (24.63)	9 (18.75)
O-IIb	7 (5.22)	2 (4.17)
O-IIc	9 (6.72)	0 (0)
0-IIa+O-IIc	47 (35.07)	22 (45.83)
0-Hc+O-IIa	18 (13.43)	10 (20.83)
0-IIb+O-Hc	12 (8.96)	3 (6.25)
O-III	0 (0)	1 (2%)
Differentiation status, Number (%)		
Differentiated	124 (92.54)	41 (85.42)
Undifferentiated	2 (1.49)	1 (2.08)
Mixed	8 (5.97)	6 (12.5)
Depth of tumor, Number (%)		
Tla	128 (95.52)	44 (91.67)
Tlb	6 (8.96)	4 (8.33)
Tla, mucosa; Tlb, submucosa.		

Comparison between the performance of DCNN and endoscopists

Eight endoscopists were selected from two hospitals and divided into the primary group and the expert group. Junior endoscopists (JE) with 2 years of operation experience and less than 1000 cases of EGD operation experience, respectively. Senior endoscopists(SE) have more than 10 years of endoscopic diagnosis and treatment experience, and each of them independently completed at least 80 cases of EGC ESD treatment; the images of the test set are arranged in a random order. Endoscopists individually read the images from the test set and recorded the time required to read the images. At the same time, DCNN recognizes the test set images and records the results.

Outcome measures and data statistics

The DCNN showed a 0–100% continuous variable number, which represented a probability score for gastric cancer in each image. Definition of correct answer, for EGC: the correct marking is the red rectangle (according to the results of the ESD postoperative pathology), the yellow rectangle is the DCNN marking and the blue rectangle is the endoscopists marking. When the yellow and red marking overlap is more than 50%, or blue and red marking overlap is greater than 50% is correct (Figure 1A); Non-cancerous lesions: The yellow rectangle box is not displayed and the word "cancer" is not displayed.95% confidence intervals (95% CI) using the modified Wald method: Agresti and Coull (The American Statistician. 52:119-126, 1998) Two-tailed unpaired Student's T-test (chi-square test)
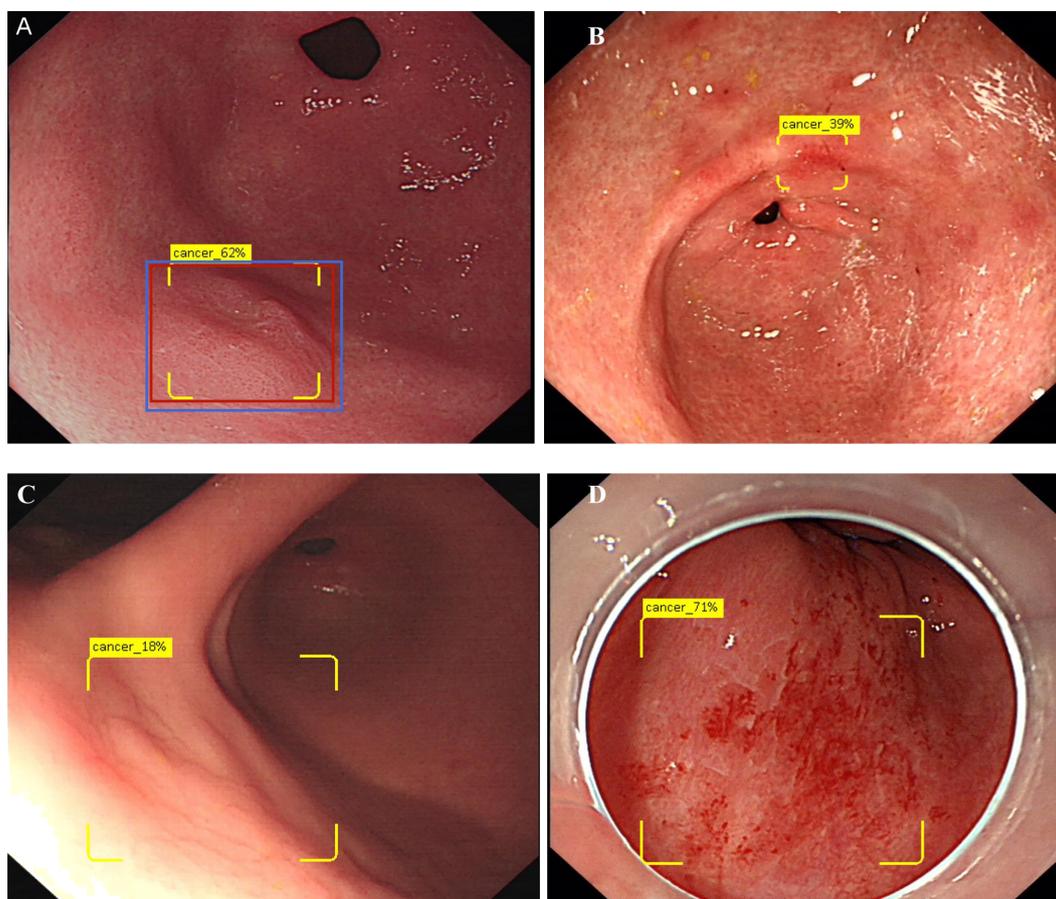
FIGURE 1
Show the concept of correctly identifies and diagram of DCNN producing false positives. **(A)** Show 0-IIa lesion in the anterior wall of gastric antrum, the red rectangle is the correct marking, the yellow box is the DCNN marking, and the blue rectangle is the endoscopist's marking. Only when the overlap reaches 50% or more, the diagnosis is correct. The yellow logo shows "Cancer 62%", indicating that DCNN predicts that the probability of EGC for this lesion is 62%. **(B)** Show an image of falsely diagnosing inflammation as EGC. **(C)** Show an image of DCNN diagnosed the normal mucosa in the reflective area as EGC. **(D)** Show an image of DCNN diagnosed the bleeding mucosa in the reflective area as EGC.

was used, with a significance level of 0.05. The accuracy, sensitivity, specificity, positive and negative predictive values (PPV and NPV, respectively) were compared. Interobserver used Cohen's Kappa coefficient (Kappa value) to assess intra-observer consistency for endoscopists. SPSS 26 (IBM, Chicago, IL, USA) was used to complete all calculations.

Ethics

The study was approved by the Ethics Committee of the Lanzhou University Second Hospital (No.2022A-004).

Results

Characteristics of patients and lesions in the test data set

The characteristics of patients and lesions in the test data set are summarized in Table 1, 95.52%patients were mucosal cancer (T1a), 4.48% patients were submucosal cancer(T1b) in ITS, and 91.67%patients were mucosal cancer (T1a), 8.33% patients were submucosal cancer(T1b) in ETS. In terms of histopathological types, 124 (92.54%) patients were differentiated gastric cancer, 10(7.46%) patients were undifferentiated gastric cancer in ITS. Differentiated cancers accounted for 41 (85.42%), and undifferentiated and mixed cancers accounted for 7 (14.58%).

The cancer diameter ranged from 4mm to 48.5mm, with a median size of 15mm in ITS, and 6-40.2mm in ETS. The most common Macroscopic type was 0-IIa+IIc, accounting for 35.07% in ITS and 45.83% in ETS, respectively.

Performance of DCNN model and endoscopists for ITS and ETS

DCNN performance

The performances of DCNN model and endoscopist are summarized in Table 2. The sensitivity, specificity, accuracy, PPV and NPV of DCNN model in ITS was 88.08%,95% confidence interval (95% CI), (85.24%-90.44%); 89.05%(95% CI, 86.48%-91.19%), 88.60%(95%CI,86.74%-90.22%), 87.64% (95%CI,84.78%-90.04%),89.44%(95%CI,86.90%-91.54%), respectively.And 92.08% (95% CI,87.91%- 94.94%),92.05%(95% CI,88.40%-94.65%), 92.07%(95%CI,89.46%-94.08%), 90.2% (95%CI, 85.79%-93.38%), 93.60(95%CI,90.17%-95.92%) in ETS. The performance of DCNN model in ETS is obviously higher than that in ITS. The average time for DCNN model analysis of each image in ITS and ETS was 0.028s.

Endoscopists performance

The endoscopist's diagnostic performances are summarized in Table 2. The diagnostic time of each image was 8.05 ± 0.21s and 7.69 ± 0.25s for the JE group and SE group in ITS, and there was no significant difference in the diagnostic time between the

TABLE 2 The performances of DCNN and endoscopist in internal test set and external test set.

Competitors	Sensitivity, % (95%CI)	Specificity, % (95%CI)	Accuracy, % (95%CI)	PPV, % (95%CI)	NPV, % (95% CI)	DT(s)
Internal test set						
DCNN	88.08	89.05	88.60	87.64	89.44	0.028s
	(85.24-90.44)	(86.48-91.19)	(86.74-90.22)	(84.78-90.04)	(86.90-91.54)	
JE group	69.54[a]	85.69	78.12[s]	81.08	76.13	8.05±0.21[a]
	(66.88-72.07)	(83.74-87.45)	(75.78-80.30)	(78.58-83.35)	(73.94-78.20)	
SE group	89.57[a]	90.00	89.73[a]	88.76	90.73	7.69±0.25[a]
	(87.71-91.17)	(88.29-91.48)	(86.79-92.08)	(86.86-90.42)	(89.07-92.16)	
External test set						
DCNN model	92.08	92.05	92.07	90.2	93.60	0.028s
	87.91-94.94	88.40-94.65	89.46-94.08	85.79-93.38	90.17-95.92	
JE group	70.42[a]	82.95	77.40[a]	76.64	77.92	7.98±0.19[a]
	66.18-74.33	79.73-85.74	74.81-79.79	72.47-80.36	74.54-80.96	
SE group	90[b]	90.23	90.13[c]	87.98	91.91	7.50±0.23[a]
	86.97-92.39	87.59-92.36	88.20-91.77	84.79-90.58	89.41-93.86	

a: P<0.01; b: P=0.02; c: P > 0.05. JE(juniorendoscopists); SE(seniorendoscopists); PPV (positive predictive value); NPV(negative predictive value); DT(diagnostic time).

two, respectively. In ITS,the sensitivity, specificity, accuracy, PPV and NPV of JE group were as follows: 69.54%(95% CI,66.88%-72.07%) 85.69%(95%CI, 83.74%-87.45%) 78.12% (95%CI, 75.78%-80.30%),81.08%(95%CI, 78.58%-83.35%) 76.13%(95%CI, 73.94%-78.20%); SE group has a better performance, 89.57(95%CI, 87.71%-91.17%) of sensitivity,90.00%(95%CI, 88.29%-91.48%)of specificity,89.73% (95%CI,86.79%-92.08%) for accuracy,88.76% (95%CI, 86.86%-90.42%)for PPV and 90.73%(95%CI, 89.07%-92.16%) for NPV. The sensitivity and accuracy of the SE group were significantly higher than those of the JE group(P<0.01), The sensitivity of SE group was higher (by20.03%95%CI, 17.03%-23.42%,P<0.01) than JE group, and the accuracy of SE group was higher (by11.61%,95%CI, 10%-13.51%,P<0.01) than JE group. In the ETS, the diagnostic time for each image in the JE and SE groups was 7.98 ± 0.19s and 7.50 ± 0.23s, respectively, and there was no significant difference in diagnostic time between the two group, the performance of SE group was significantly higher than that of JE group, the sensitivity of SE group was higher (by19.58%,95%CI, 15.04%-25.09%,P<0.01) than that of JE group, and the accuracy of SE group was higher (by12.73%,95%CI, 10.17%-15.81%, P<0.01) than that of JE group. In terms of the ITS and the ETS, the diagnostic efficacy of endoscopists in the ETS was higher than that in the ITS.

In terms of diagnostic consistency, in the ITS, the DCNN model and endoscopist's pairwise Kappa values ranged from 0.765 to 0.913, while the endoscopist's diagnostic Kappa values ranged from 0.735 to 0.959. The diagnostic consistency was reasonable. The mean value of Kappa between DCNN model and endoscopist was 0.8794, JE-1 was 0.8424, JE-2 was 0.871, SE-1 was 0.8868, and SE-2 was 0.878. In the ETS, the Kappa values of DCNN model and endoscopists ranged from 0.676 to 0.981, while the diagnostic Kappa values of endoscopist ranged from 0.682 to 0.950. The mean value of Kappa between DCNN model and endoscopist's was 0.8638, JE-3 was 0.8106, JE-4 was 0.8418, SE-3 was 0.8832, and SE-4 was 0.8686.

Comparison of DCNN and endoscopist performance

The receiver operating characteristic (ROC) curve of DCNN model and endoscopist's diagnostic effectiveness is shown in Figure 2. In ITS, the area under ROC curve (AUC) of the DCNN model, JE group and SE group were 0.8857 (95%CI,0.8655-0.9058),0.7710 (95%CI,0.7443-0.7978) and 0.8890(95% CI,0.8690-0.9091),respectively. In ETS, AUC of the DCNN model, JE group and SE group were 0.9207 (95%CI,0.9020-0.9394), 0.7668(95%CI,0.7372-0.7964) and 0.9012 (95% CI,0.8805-0.9218), respectively. DCNN model was significantly faster than all endoscopists in test sets. The sensitivity of the DCNN model was 18.54% (95%CI, 15.64%-21.84%,P<0.01) higher than that of the JE group and 0.33%(95%CI, 0.06%-1.05%,P>0.05) lower than that of the SE group in the ITS. In the ETS, it was 21.67% (95%CI, 16.90%-27.32%, P<0.01) higher than
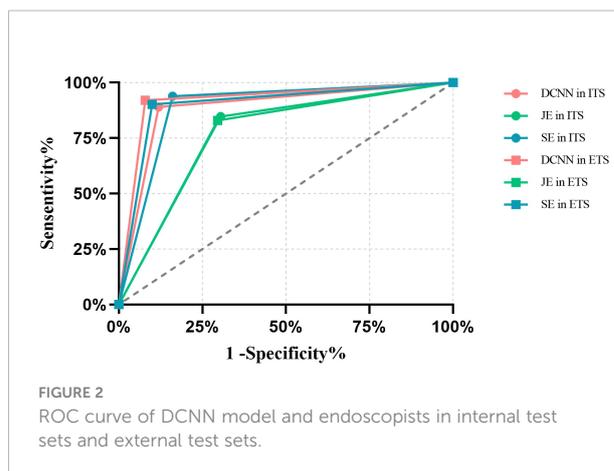
ROC curve of DCNN model and endoscopists in internal test sets and external test sets.

that in the JE group and 2.08%(95%CI, 0.75%-4.92%, P>0.05) higher than that in the SE group. In terms of accuracy, DCNN model was 10.47%(95%CI, 8.91-12.27%,P<0.05) higher than that of JE group, and 1.16% (95%CI, 0.69%-1.93%,P > 0.05) lower than that of SE group in the ITS; In the ETS, it was 14.58%(95% CI, 11.84%-17.81%, P<0.01) higher than JE group and 1.94% (95%CI,1.25%-2.96%,P > 0.05)higher than SE group.

Cause of false positives and false negatives

In order to further analyze the causes of false positives and false negatives produced by DCNN model and endoscopists, we summarized it in Tables 3, 4. The first cause for false positive of DCNN model is Gastritis (redness, atrophy, intestinal metaplasia)(44% and 62.5%,respectively),which were also the most reasons for endoscopists(59.29% and 68.52%).Mucus (10.67%) was the secondary cause of in the ITS, while ulcer (12.5%) was the secondary cause in the ETS, which we found that the surface appearance of ulcers were very similar to that of gastric cancer. The third false-positive factor was folding and foam (9.33% in the ITS) and blood (8.33% in the ETS). For endoscopists, ulcer was the second reason (11.43% and 18.52%). However, compared with the DCNN model, endoscopists rarely mistake mucus, foam, and folding for EGC.

Table 4 summarizes the causes of false negatives, the first reason for the false negative in DCNN model was that the diameter of the lesions was less than 10mm (38.88% and 21.05%, respectively). And the 32 images were from 14 patients respectively, DCNN model could identify some of the images in these cases, however the long shooting distance combined with the small diameter of the lesions was the biggest reason for the error recognition of DCNN model. The second factor was visual angle (25% in the ITS), distance and ulcers (15.79% in the ETS, respectively). Different shooting angles resulted in incomplete identification of multiple images of the same

TABLE 3   Details of DCNN model and false positive images of endoscopists.

Cause	Internal test set		External test set	
	DCNN, n (%)	Endoscopists, Total n	DCNN,n	Endoscopists, Total n
		(%)	(%)	(%)
Total number	75	280	24	162
Gastritis (redness, atrophy, intestinal metaplasia)	33 (44)	166 (59.29)	15 (62.5)	111 (68.52)
Mucus	8 (10.67)	2 (0.71)	1 (4.17)	4 (2.47)
Fold	7 (9.33)	11 (3.93)	0 (0)	7 (4.32)
Foam	7 (9.33)	4 (1.43)	1 (4.17)	0 (0)
Halation	5 (6.67)	5 (1.79)	0 (0)	0 (0)
Blood	4 (5.33)	8 (2.86)	2 (8.33)	4 (2.47)
blood vessel	4 (5.33)	0 (0)	0 (0)	0 (0)
Ulcer	3 (4)	32 (11.43)	3 (12.5)	30 (18.52)
Xanthoma	2 (2.67)	11(3.93)	1 (4.17)	3 (1.85)
Normal anatomical structure (cardia, pylorus, angulus)	1 (1.33)	0 (0)	0 (0)	0 (0)
Hyperplastic polyp	1 (1.33)	30(10.71)	0 (0)	2 (1.23)
Scar	0 (0)	11(3.93)	1 (4.17)	1 (0.62)

lesion, some lesions in these unrecognized images were relatively flat (type 0-IIb), some of the light increased as the shooting angle changed. The third factor of false negative was distant (16.67% in ITS), tangential line and inflammation-like(10.53%, respectively). For endoscopists, the biggest factor of false negative is inflammation-like (32.74% and 56.84%). Examples of false positive and false negative by DCNN model was shown in Figures 1B–D, 3A–D.

Discussion

In the world, only Japan and South Korea have relatively high diagnosis rate of EGC, while European and American countries have not carried out large-scale endoscopic screening of EGC. China has a large population, and the incidence of gastric cancer accounts for 1/4 of the world, but the diagnosis rate of EGC is only 10% (3, 4), it is far behind Japan and South

TABLE 4   Details of false negative images by DCNN model and endoscopists.

| Cause	Internal test set		External test set	
	DCNN, n (%)	Endoscopists, total n (%)	DCNN, n (%)	Endoscopists, totaln (%)
Total number	72	562	19	190
Small (< 10mm)	28 (38.88)	112 (19.93)	4 (21.05)	16 (8.42)
visual angle	18 (25)	50 (8.90)	1 (5.3)	2 (1.05)
Distant	12 (16.67)	90 (16.01)	3 (15.79)	12 (6.32)
Tangential Line	5 (6.94)	12 (2.14)	2 (10.53)	4 (2.1)
Ulcer	4 (5.56)	91 (16.19)	3 (15.79)	28 (14.74)
Adenoma-like	2 (2.78)	16 (2.85)	1 (5.3)	4 (2.1)
Inflammation-like	2 (2.78)	184 (32.74)	2 (10.53)	108 (56.84)
Blood	1 (1.39)	2 (0.36)	1 (5.3)	6 (3.16)
Scar-like	0 (0)	5 (0.89)	0 (0)	5 (2.63)
Others	0 (0)	0 (0)	2 (10.53)	5 (2.63)
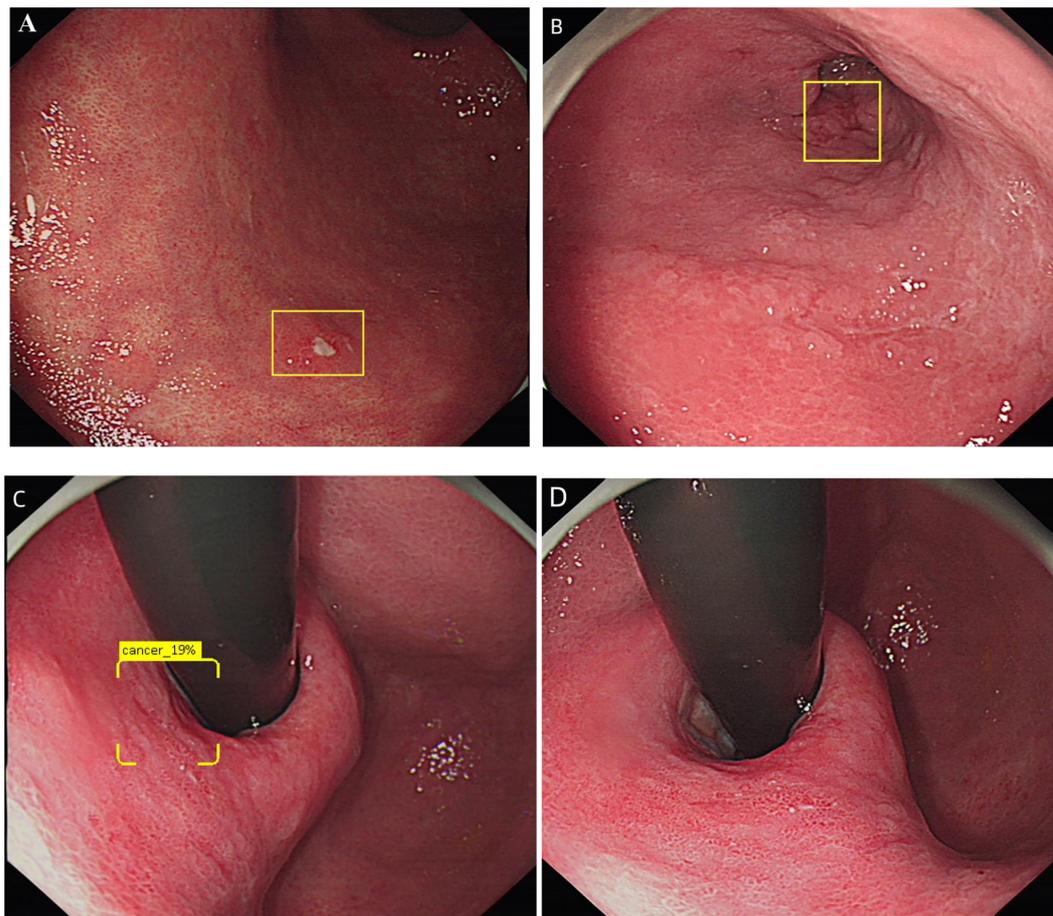
FIGURE 3
Diagram of DCNN producing false negatives. **(A)** Show an image that the lesions were too small to be recognized by DCNN. **(B)** Show an image of the lesions were too far to be recognized by DCNN. **(C, D)** Shows images taken from different shooting angles of the same lesion. **(C)** was effectively recognized as EGC by DCNN, while **(D)** was not recognized by DCNN.

Korea. The difficulty in treatment of EGC lies in early detection. EGD is the only effective tool to identify EGC. Although standardized training of EGD can improve the diagnosis rate, However, the time of training curve is long, and the scope of standardized training is limited (20, 21), in addition, there is a serious shortage of specialized endoscopists in China, the overall level of diagnosis rate in EGC is low, and it is difficult to improve the diagnosis rate of EGC in a short period of time. Therefore, how to quickly shorten endoscopists training time and improve the level of diagnosis rate of EGC is an urgently problem for us to solve.

Although various image enhancement techniques have been developed and applied, white light imaging(WLI) is the first step in standard EGD (16), The use of image enhancement technology is considered only after suspicious lesion under WLI. It has been reported that the sensitivity of WLI for EGC is 33%-75% (22), On the other hand, diagnosis depends on the experience and subjective awareness of the EGD operator (17).

Therefore, through the training of images, we developed the DCNN system to assist the diagnosis of EGC in WLI. The sensitivity of DCNN model is 88.08% and 92.08%, which was significantly higher than that of all endoscopists, and its recognition time was significantly shorter than that of all endoscopists.

We found that ulcers (11.43% and 18.52%) were an important cause of false positives produced by endoscopists, as well as a cause of missed diagnosis (16.19% and 14.74%), which means that endoscopists have difficulty in distinguishing ulcers from EGC. Ulcers were included as negative controls in our DCNN model training. Therefore, compared with endoscopists, the proportion of false positives in DCNN model due to ulcers was lower than that of endoscopists (4%-12.5% vs.11.43%-18.52%). This means that our DCNN model can help endoscopists reduce the incidence of such errors. Of course, our DCNN model also has a certain false negative rate. Among them, 38.88% and 21.05% are lesions smaller than 10mm.

Considering that it is difficult for even experienced endoscopists to diagnose small lesions, the development time of intratumoral cancer is 2-3 years, we speculate that this limitation can be confirmed by annual upper gastrointestinal endoscopy and biopsy (23), and for the update iteration of DCNN, we will also increase the training of small lesions. 25% and 5.3% false negatives are difficult to identify because of different angles of view. However, DCNN can identify at least two images of a patient from different angles, therefore, we speculate that the retention of suspicious lesions from multiple locations and angles can make up for this defect of DCNN. The lesion distance is the third major reason leading to false negative of DCNN. Although the morphology of these lesions is prominent, but the images were taken as a far distance, and the color, texture and other features of the lesions were not obvious, therefore it is easy to be ignored by the DCNN. Another rare reason for false negatives in DCNN is that lesions are similar to inflammatory changes, but they are the most common cause for endoscopists. This finding suggests that endoscopists may misdiagnose inflammation, but DCNN does not miss lesions (2.78% vs. 32.74% in ITS, 10.53% vs.56.84% in ETS).

The most common causes of false positives are mucosal redness, atrophy, and intestinal degeneration. Even experienced endoscopists can hardly distinguish these lesions by a single WLI without magnifying endoscopy. 29.33% and 8.34% of false positives are due to mucus, foam and folding. DCNN is more likely to be affected by the above factors, while endoscopists are less affected by these factors, which means that DCNN may be more affected by the background in the stomach in the actual application process. However, as demonstrated by Mori et al. (24) these limitations can be reduced by mucosal irrigation, the use of antifoaming agents and adequate gas injection, as well as the application of a large number of images for training in the real process of EGD.

For specificity and PPV, Although SE performed better on specificity and PPV in ITS, the DCNN model was higher than that of JE, our image of training set goes through carefully selected, excluded those poor quality of the image images, containing mucus, bubble, folding and dizzy light images, We believe that if we strengthen the training of false positive images, these problems will be solved (25). DCNN is more sensitive and can identify more EGCs than experienced endoscopists, especially for JE. In addition, the sensitivity and PPV of expert endoscopists are significantly higher than those of JE, so DCNN may be more helpful to those endoscopists with limited experience.

We reviewed the relevant literature and found that some AI have a sensitivity of up to 90% (26–28). However, the control groups in these studies only included normal or chronic gastritis, not ulcerative lesions that endoscopists are more likely to confuse. and there studies focused on the sensitivity of detecting gastric cancer as a whole, including advanced gastric cancer, and were not compared with endoscopists. Moreover, most studies did not analyze the causes of false positive and false

negative for endoscopists and DCNN, so as to carry out targeted strengthening training. The sensitivity found in this study appears to be lower in ITS than those studies for the following reasons: first, only EGC images were included in this study, and intermediate and advanced gastric cancer were not included. Second, sensitivity was calculated per lesion, not per image in those studies, that is, if at least one image of gastric cancer is recognized in multiple images of the same lesion, the diagnosis is considered correct. Third, in these studies, the concept of correct diagnosis is that if the image area identified by DCNN overlaps slightly with that of EGC, it is considered correct. However, in our study, only the image range identified by DCNN overlapped by more than 50% with the image range marked by endoscopist was considered correct, while occasionally marked or not marked enough, we considered it incorrect. Recently, A multicenter study reported (29) that DCNN assisted diagnosis of upper gastrointestinal tumors, including gastric cancer shows diagnostic accuracy of DCNN was up to 90%, and the sensitivity was comparable with that of expert endoscopists. However, in this study, the rate of advanced gastric cancer was relatively higher, while the early gastric cancer was only 18.6%. Our research aims to find more EGC that endoscopists are prone to misdiagnose.

In terms of stability evaluation, SE group with higher diagnostic accuracy has better diagnostic consistency than JE group. According to the general guidelines of ICC standard (30), there are considerable differences in the diagnosis consistency among endoscopists, which is not clearly related to professional knowledge and experience. Due to the subjective interpretation of the characteristics of EGC and the different learning curve in the diagnosis of EGC, objective diagnosis is very necessary (9). The DCNN system achieves perfect observer protocol (Kappa 1.0) without interference of subjective judgment. The EGC detection system based on the DCNN has sufficient and consistent diagnostic performance, eliminating some diagnostic subjectivity. Moreover, the DCNN system is very helpful for JE, this is consistent with the research of Ikenoyama et al. (31) DCNN may be a powerful tool to assist endoscopists, especially JE in detecting EGC. The shorter screening time and fatigue free DCNN may enable rapid surveillance of EGC. More importantly, the diagnosis of EGC by DCNN can be fully automated and online, which may facilitate the development of telemedicine and thereby alleviate the problem of a shortage of endoscopists.

Yet, the study has several limitations, first of all, our DCNN only trained WLI images, it can provides the first step of EGC detection, which is also the most important step, but it did not trained narrow-band images (NBI) and the magnifying endoscopy (ME). However. in past research report, endoscopic image enhancement is rarely used, unless there is a suspicious lesion found in the WLI (32). In addition, a multicenter study showed no significant difference in the diagnostic efficacy of nonamplified NBI and WLI in EGC (33). Moreover, in reality,

unless use the NBI routinely in esophageal observation, It is generally not used in endoscopic examination unless suspicious lesions are found under WLI (32). Secondly, we only use Olympus 260 or 290 series gastroscope system, without other brands of endoscopes such as Fuji Endoscope, this may reduce the efficiency of DCNN. Third, in the control group of gastric cancer detection dataset, we eliminated most of the images containing mucus and halo, and the diagnostic sensitivity of DCNN may be lower in the real world. Fourth, static images are used in the training and testing sets of this study, and video images can improve the performance and present the real scene, we plan to use video as a validation set in the future, which will be used as another separate study. Fifth, in this study, DCNN model missed diagnosis of small lesions, increasing the number of training of small lesions and flat lesions, as well as the number of negative control images will help improve the diagnostic efficiency of the model.

In conclusion, we constructed an assist EGC detection system based on DCNN and compared the diagnostic ability of DCNN and endoscopists. It has excellent diagnostic sensitivity, fast diagnostic characteristics, achieved a perfect observer protocol. It can help endoscopists (especially JE) to find more EGC. We believe that DCNN will contribute to the overall improvement of the diagnosis rate of EGC, and serve as an assisting work to help improve the diagnosis rate of EGC.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Ethics statement

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

JF designed the experiment and drafted this article. SY and YZ were responsible for the collection of training set images. LQ was responsible for the submission of ethical approval documents. YF nd YY were responsible for the annotation of the images. LG, XL, WL and PW were responsible for the recognition of the test set images. YB and LZ were responsible for the algorithm training of DCNN. XH was responsible for the project coordination. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.1021625/full#supplementary-material

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: Cancer J Clin* (2021) 71(3):209–49. doi: 10.3322/caac.21660

2. Katai H, Ishikawa T, Akazawa K, Isobe Y, Miyashiro I, Oda I, et al. Five-year survival analysis of surgically resected gastric cancer cases in Japan: a retrospective analysis of more than 100,000 patients from the nationwide registry of the Japanese gastric cancer association (2001-2007). *Gastric Cancer.* (2018) 21(1):144–54. doi: 10.1007/s10120-017-0716-7

3. Baptista V, Singh A, Wassef W. Early gastric cancer: an update on endoscopic management. *Curr Opin gastroenterology.* (2012) 28(6):629–35. doi: 10.1097/MOG.0b013e328358e5b5

4. Bu Z, Ji J. Controversies in the diagnosis and management of early gastric cancer. *Chin J Cancer Res = Chung-kuo yen cheng yen chiu.* (2013) 25(3):263–6. doi: 10.3978/j.issn.1000-9604.2013.06.15

5. Asaka M, Mabe K. Strategies for eliminating death from gastric cancer in Japan. *Proc Japan Acad Ser B Phys Biol Sci* (2014) 90(7):251–8. doi: 10.2183/pjab.90.251. Cited in: Pubmed

6. Amin A, Gilmour H, Graham L, Paterson-Brown S, Terrace J, Crofts TJ. Gastric adenocarcinoma missed at endoscopy. *J R Coll Surgeons Edinburgh.* (2002) 47(5):681–4. doi: 10.1055/s-2004-825853

7. Hosokawa O, Tsuda S, Kidani E, Watanabe K, Tanigawa Y, Shirasaki S, et al. Diagnosis of gastric cancer up to three years after negative upper gastrointestinal endoscopy. *Endoscopy.* (1998) 30(8):669–74. doi: 10.1055/s-2007-1001386

8. Menon S, Trudgill N. How commonly is upper gastrointestinal cancer missed at endoscopy? A meta-analysis. *Endoscopy Int Open* (2014) 2(2):E46–50. doi: 10.1055/s-0034-1365524

9. Yalamarthi S, Witherspoon P, McCole D, Auld CD. Missed diagnoses in patients with upper gastrointestinal cancers. *Endoscopy.* (2004) 36(10):874–9. doi: 10.1055/s-2004-825853

10. Suvakovic Z, Bramble MG, Jones R, Wilson C, Idle N, Ryott J. Improving the detection rate of early gastric cancer requires more than open access gastroscopy: a five year study. *Gut.* (1997) 41(3):308–13. doi: 10.1136/gut.41.3.308

11. Voutilainen ME, Juhola MT. Evaluation of the diagnostic accuracy of gastroscopy to detect gastric tumours: clinicopathological features and prognosis of patients with gastric cancer missed on endoscopy. *Eur J Gastroenterol hepatology.* (2005) 17(12):1345–9. doi: 10.1097/00042737-200512000-00013

12. Raftopoulos SC, Segarajasingam DS, Burke V, Ee HC, Yusoff IF. A cohort study of missed and new cancers after esophagogastroduodenoscopy. *Am J gastroenterology.* (2010) 105(6):1292–7. doi: 10.1038/ajg.2009.736

13. Vradelis S, Maynard N, Warren BF, Keshav S, Travis SP. Quality control in upper gastrointestinal endoscopy: detection rates of gastric cancer in Oxford 2005-2008. *Postgraduate Med J* (2011) 87(1027):335–9. doi: 10.1136/pgmj.2010.101832

14. Hosokawa O, Hattori M, Douden K, Hayashi H, Ohta K, Kaizaki Y. Difference in accuracy between gastroscopy and colonoscopy for detection of cancer. *Hepato-gastroenterology.* (2007) 54(74):442–4. doi: 10.1055/s-2001-13685

15. Rutter MD, Senore C, Bisschops R, Domagk D, Valori R, Kaminski MF, et al. The European society of gastrointestinal endoscopy quality improvement initiative: developing performance measures. *United Eur Gastroenterol J* (2016) 4(1):30–41. doi: 10.1177/2050640615624631

16. Yao K, Uedo N, Muto M, Ishikawa H. Development of an e-learning system for teaching endoscopists how to diagnose early gastric cancer: basic principles for improving early detection. *Gastric Cancer.* (2017) 20(Suppl 1):28–38. doi: 10.1007/s10120-016-0680-7

17. Scaffidi MA, Grover SC, Carnahan H, Khan R, Amadio JM, Yu JJ, et al. Impact of experience on self-assessment accuracy of clinical colonoscopy competence. *Gastrointestinal endoscopy.* (2018) 87(3):827–36. doi: 10.1016/j.gie.2017.10.040

18. Wang KW, Dong M. Potential applications of artificial intelligence in colorectal polyps and cancer: Recent advances and prospects. *World J gastroenterology.* (2020) 14 26(34):5090–100. doi: 10.3748/wjg.v26.i34.5090

19. Xia J, Xia T, Pan J, Gao F, Wang S, Qian YY, et al. Use of artificial intelligence for detection of gastric lesions by magnetically controlled capsule endoscopy. *Gastrointestinal endoscopy.* (2021) 93(1):133–139.e4. doi: 10.1016/j.gie.2020.05.027

20. Yamazato T, Oyama T, Yoshida T, Baba Y, Yamanouchi K, Ishii Y, et al. Two years' intensive training in endoscopic diagnosis facilitates detection of early gastric cancer. *Internal Med (Tokyo Japan).* (2012) 51(12):1461–5. doi: 10.2169/internalmedicine.51.7414

21. Zhang Q, Chen ZY, Chen CD, Liu T, Tang XW, Ren YT, et al. Training in early gastric cancer diagnosis improves the detection rate of early gastric cancer: An observational study in China. *Medicine.* (2015) 94(2):e384. doi: 10.1097/md.0000000000000384

22. Ezoe Y, Muto M, Uedo N, Doyama H, Yao K, Oda I, et al. Magnifying narrowband imaging is more accurate than conventional white-light imaging in diagnosis of gastric mucosal cancer. *Gastroenterology.* (2011) 141(6):2017–2025.e3. doi: 10.1053/j.gastro.2011.08.007

23. Fujita S. Biology of early gastric carcinoma. *Pathology Res practice.* (1978) 163(2):297–309. doi: 10.1016/s0344-0338(78)80028-4

24. Mori Y, Kudo SE, Mohmed HEN, Misawa M, Ogata N, Itoh H, et al. Artificial intelligence and upper gastrointestinal endoscopy: Current status and future perspective. *Dig Endosc.* (2019) 31(4):378–88. doi: 10.1111/den.13317

25. Gotoda T, Uedo N, Yoshinaga S, Tanuma T, Morita Y, Doyama H, et al. Basic principles and practice of gastric cancer screening using high-definition white-light gastroscopy: Eyes can only see what the brain knows. *Dig Endosc.* (2016) 28 Suppl 1:2–15. doi: 10.1111/den.12623

26. Hirasawa T, Aoyama K, Tanimoto T, Ishihara S, Shichijo S, Ozawa T, et al. Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images. *Gastric Cancer.* (2018) 21(4):653–60. doi: 10.1007/s10120-018-0793-2

27. Luo H, Xu G, Li C, He L, Luo L, Wang Z, et al. Real-time artificial intelligence for detection of upper gastrointestinal cancer by endoscopy: a multicentre, case-control, diagnostic study. *Lancet Oncol* (2019) 20(12):1645–54. doi: 10.1016/s1470-2045(19)30637-0

28. Wu L, Zhou W, Wan X, Zhang J, Shen L, Hu S, et al. A deep neural network improves endoscopic detection of early gastric cancer without blind spots. *Endoscopy.* (2019) 51(6):522–31. doi: 10.1055/a-0855-3532

29. Luo H, Xu G, Li C, He L, Luo L, Wang Z, et al. Real-time artificial intelligence for detection of upper gastrointestinal cancer by endoscopy: A multicentre, case-control, diagnostic study. *Lancet Oncol* (2019) 20(12):1645–54. doi: 10.1016/s1470-2045(19)30637-0

30. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J chiropractic Med* (2016) 15(2):155–63. doi: 10.1016/j.jcm.2016.02.012

31. Ikenoyama Y, Hirasawa T, Ishioka M, Namikawa K, Yoshimizu S, Horiuchi Y, et al. Detecting early gastric cancer: Comparison between the diagnostic ability of convolutional neural networks and endoscopists. *Dig Endosc.* (2021) 33(1):141–50. doi: 10.1111/den.13688

32. Uedo N, Gotoda T, Yoshinaga S, Tanuma T, Morita Y, Doyama H, et al. Differences in routine esophagogastroduodenoscopy between Japanese and international facilities: A questionnaire survey. *Dig Endosc.* (2016) 28 Suppl 1:16–24. doi: 10.1111/den.12629

33. Ang TL, Pittayanon R, Lau JY, Rerknimitr R, Ho SH, Singh R, et al. A multicentre randomized comparison between high-definition white light endoscopy and narrow band imaging for detection of gastric lesions. *Eur J Gastroenterol hepatology.* (2015) 27(12):1473–8. doi: 10.1097/meg.0000000000000478