Zou et al. BMC Gastroenterology (2024) 24:335
https://doi.org/10.1186/s12876-024-03389-3

BMC Gastroenterology

RESEARCH

Open Access

Development and application of an artificial intelligence-assisted endoscopy system for diagnosis of *Helicobacter pylori* infection: a multicenter randomized controlled study

Pei-Ying Zou[1], Jian-Ru Zhu[1], Zhe Zhao[1], Hao Mei[1], Jing-Tao Zhao[1], Wen-Jing Sun[2], Guo-Hua Wang[3], Dong-Feng Chen[1], Li-Lin Fan[4] and Chun-Hui Lan[1*]

Abstract

Background The early diagnosis and treatment of *Heliobacter pylori* (*H.pylori*) gastrointestinal infection provide significant benefits to patients. We constructed a convolutional neural network (CNN) model based on an endoscopic system to diagnose *H. pylori* infection, and then examined the potential benefit of this model to endoscopists in their diagnosis of *H. pylori* infection.

Materials and methods A CNN neural network system for endoscopic diagnosis of *H.pylori* infection was established by collecting 7377 endoscopic images from 639 patients. The accuracy, sensitivity, and specificity were determined. Then, a randomized controlled study was used to compare the accuracy of diagnosis of *H. pylori* infection by endoscopists who were assisted or unassisted by this CNN model.

Results The deep CNN model for diagnosis of *H. pylori* infection had an accuracy of 89.6%, a sensitivity of 90.9%, and a specificity of 88.9%. Relative to the group of endoscopists unassisted by AI, the AI-assisted group had better accuracy (92.8% [194/209; 95%CI: 89.3%, 96.4%] *vs.* 75.6% [158/209; 95%CI: 69.7%, 81.5%]), sensitivity (91.8% [67/73; 95%CI: 85.3%, 98.2%] *vs.* 78.6% [44/56; 95%CI: 67.5%, 89.7%]), and specificity (93.4% [127/136; 95%CI: 89.2%, 97.6%] *vs.* 74.5% [114/153; 95%CI: 67.5%, 81.5%]). All of these differences were statistically significant ($P < 0.05$).

Conclusion Our AI-assisted system for diagnosis of *H. pylori* infection has significant ability for diagnostic, and can improve the accuracy of endoscopists in gastroscopic diagnosis.

Trial registration This study was approved by the Ethics Committee of Daping Hospital (10/07/2020) (No.89,2020) and was registered with the Chinese Clinical Trial Registration Center (02/09/2020) (www.chictr.org.cn; registration number: ChiCTR2000037801).

Keywords *Helicobacter pylori*, Endoscopy, Artificial intelligence, Convolutional neural network

*Correspondence:
Chun-Hui Lan
lanchunhui@tmmu.edu.cn
Full list of author information is available at the end of the article

Introduction

Helicobacter pylori is a microaerobic Gram-negative bacillus that can cause gastrointestinal diseases, such as chronic active gastritis, peptic ulcer, gastric mucosa-associated lymphoid tissue lymphoma, and gastric cancer. About 50% of people worldwide have *H. pylori* infections. Early diagnosis of *H. pylori* infection and eradication are important for treating chronic active gastritis and reducing the recurrence of peptic ulcer and the occurrence of gastric cancer [1–3].

Invasive and non-invasive methods are currently used to diagnose *H. pylori* infection. The main invasive methods are the Rapid Urease Test (RUT), histopathological diagnosis, endoscopic diagnosis, and molecular analysis; the main non-invasive methods are the Urea Breath Test (UBT), Stool Antigen Test, and serological examination [4–6]. In China, the cost of an endoscopic examination is relatively low and this procedure is widely accepted by patients. China also has a high rate of infection by *H. pylori* and a high incidence of gastric cancer, so if two tests can be combined and completed simultaneously using gastroscopy, *H. pylori* infection can be treated in a timely manner, precancerous lesions of gastric cancer and early gastric cancer can be detected, and medical costs can be reduced. Although invasive methods using gastroscopy provide these benefits, there are also has some shortcomings, in that the RUT is susceptible to false negatives depending on the load and focal distribution of *H. pylori*. However, histopathological and molecular biological tests cannot provide real-time results.

After *H. pylori* infection, endoscopy can identify the unique characteristics of the gastric mucosa, and enable the diagnosis of *H. pylori* infection [7, 8]. Gastric mucosa that is positive for *H. pylori* typically has diffuse or spotty redness, mucosal swelling, with or without white turbid mucus, atrophy, intestinal metaplasia, disappearance of regular arrangement of collecting venules (RAC), nodular changes, xanthoma, and hyperplastic polyps. Gastric mucosa that is negative for *H. pylori* typically has a RAC, and may have fundic gland polyposis, a red streak in the gastric antrum and gastric body, and attachment of a former bleeding spot. After *H. pylori* eradication, there are often localized red areas of various sizes on white mucosa, and redness after erosion healing appears as patchy redness. After *H.pylori* eradication, there are often appear as patchy redness.

A previous study that used receiver operating characteristic (ROC) analysis found that use of conventional gastroscopy with white light imaging (WLI) for the diagnosis of *H. pylori* infection had an area under the curve (AUC) of 0.783, with a total coincidence rate of 78.46%, a sensitivity of 77.44%, a specificity of 79.31%, a positive predictive value (PPV) of 75.74%, and a negative predictive value (NPV) of 80.83% [9]. This previous study was based on the endoscopic observations of 2 experienced endoscopists (each with more than 5 years of endoscopy experience and examination of more than 5000 cases using gastroscopy). This previous study also reported that the accuracy of diagnosis depended on the experience of the endoscopist, and that diagnosis was time-consuming and adversely affected by fatigue. Therefore, objective, accurate, fast, and feasible endoscopic diagnosis of *H. pylori* is needed in clinical practice. Several studies reported that artificial intelligence (AI) provided accurate and standardized endoscopic diagnosis of *H. pylori* infection [9–12]. Convolutional neural networks (CNNs) are commonly used for analysis of medical images. CNNs have been widely applied in gastroenterology, and are currently the most widely used network architecture for medical image deep learning [13].

The present study established an endoscopic system for the diagnosis of *H. pylori* infection based on deep learning and applied it to a clinical setting. In particular, we compared the accuracy of gastroscopic diagnosis of *H. pylori* infection by endoscopists who were assisted by AI with diagnosis by endoscopists who were unassisted by AI, and then analyzed the possible clinical applications of this AI diagnostic system.

Methods

Study design and patients

This study consisted of a diagnostic study and a multicenter randomized controlled study. Firstly, an endoscopic system for the diagnosis of *H. pylori* infection based on a deep learning CNN model was constructed and tested. Then, the diagnostic performance of AI-assisted and AI-unassisted endoscopists were compared. The main indicators were accuracy, sensitivity and specificity, and the secondary indicators were PPV and NPV (Fig. 1).

From September 2020 to July 2021, 1258 patients were screened at two institutions (Department of Gastroenterology of Chongqing Daping Hospital, Chongqing 13th People's Hospital and Chongqing Jiulongpo District Second People's Hospital). All included patients signed written informed consent documents. The inclusion criteria were: (*i*) age of 18 to 70 years old, male or female; (*ii*) completion of two or more diagnostic tests (UBT, RUT, *H. pylori* culture, *H. pylori* histology) with consistent results; and (*iii)* more than one year after *H. pylori* treatment. The exclusion criteria were: (*i*) use of an antibiotic, bismuth, histamine H_2-receptor antagonist, proton pump inhibitor (PPI), or probiotic during the 4 weeks before testing; (*ii*) women who were planning to become pregnant, or were pregnant or breastfeeding; (*iii*) use of an adrenocorticosteroid, nonsteroidal anti-inflammatory

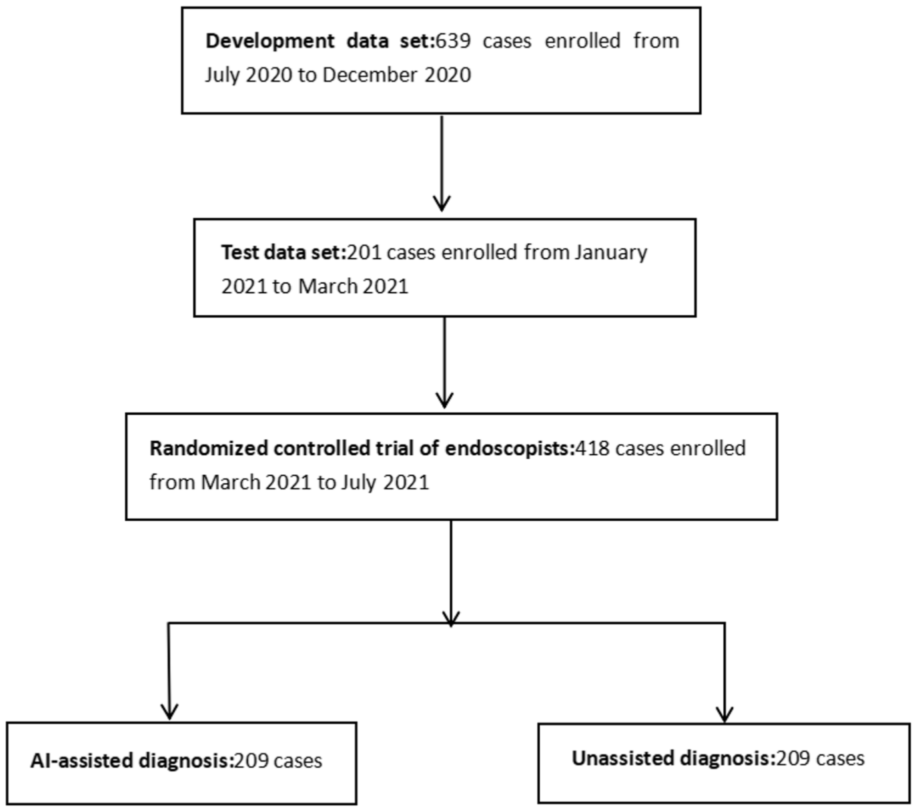

Fig. 1 Patient selection and study design

drug, or anticoagulant; (*iv*) pre-existing serious underlying disease (liver disease, cardiovascular disease, lung disease, kidney disease, metabolic disease, mental illness, malignant tumor, etc.) that made it difficult to tolerate gastroscopy; (*v*) previous history of gastric or duodenal ulcer, gastric malignant tumor, or other gastric organic lesions; (*vi*) participation in another clinical study within 3 months before participating in the present study; and (*vii*) difficulty in completing follow-up or other factors that affect compliance.

Table 1 Baseline demographic and clinical characteristics of patients in the development data set and the test data set

Characteristic	Development data	Test data	All
Number of cases	639	201	840
Number of images	7377	2080	9457
Mean age, years (SD)	45.0 ± 13.5	44.3 ± 12.2	44.8 ± 13.2
Sex, *n* (%)			
Male	289(45)	90(45)	379(45)
Female	350(55)	111(55)	461(55)
H. pylori status, *n* (%)			
Uninfected	379(59)	135(67)	514(61)
Currently infected	260(41)	66(33)	326(39)

Endoscopy equipment

Fourteen endoscopists performed the esophagogastroduodenoscopies (EGDs) using standard EGD equipment (CLV-290SL, CV-260SL, GIF-H260, GIF-H260Z, GIF-Q260J, GIF-Q260, GIF-H290, GIF-H290Z; Olympus Medical Systems).

Development data

Two endoscopists collected gastric images of 639 patients, 260 who were *H. pylori*-positive and 379 who were *H. pylori*-negative (Table 1). For each patient, there was at least one image of each of the five anatomical sites in the stomach (upper body, middle body, lower body, lesser curvature, and antrum), and there were five or more total images per patient. Representative images were selected to establish the CNN training model, and the model was then trained and constructed. Finally, 7377 images from these 639 patients were used for model learning.

Test data

To evaluate the diagnostic accuracy of the AI system, a separate set of test data was prepared. These test data consisted of 2080 images of 201 patients, 66 who were *H. pylori*-positive and 135 who were *H. pylori*-negative (Table 1). All of these patients received EGDs in the Department of Gastroenterology, Daping Hospital of Chongqing from January to March 2021. There was no overlap of patients between the development and test data sets.

Multicenter randomized controlled trial

Endoscopists were randomized to an AI-assisted group or an AI-unassisted group, each with 14 endoscopists. Three endoscopists in each group were trained in the early detection of cancer at classes in the Shanghai Jiaotong University and obtained a training certificate; the 11 other endoscopists in each group were not trained in the early detection of cancer. Each group completed endoscopy independently in AI-assisted or AI-unassisted mode, as appropriate. After the examination, the AI-assisted group received the diagnostic results in real time, and had the opportunity to make a final judgment based on these AI results. Endoscopists in the AI-unassisted group made judgments of *H. pylori* infection after examination. There was no overlap of patients at any stage. This study was approved by the Ethics Committee of Daping Hospital and was registered with the Chinese Clinical Trial Registry(www.chictr.org.cn; registration number: ChiCTR2000037801).

Training algorithm

To establish a system with high accuracy and diagnostic performance for the detection of *H. pylori* infection, EfficientNet (https://arxiv.org/pdf/1905.11946.pdf), a CNN developed by Google Research, was used. Several previously proposed dimensions of model zooming were investigated: network depth, network width, and image resolution. Previous research mostly enlarged one of these dimensions to achieve high accuracy. For example, ResNet-18 and ResNet-152 improve accuracy by increasing the network depth. Through the optimal combination of scaling these three dimensions, an EfficientNet neural network model can consider speed and accuracy.

A series of EfficientNet models were obtained by amplifying the basic model. This series of models outperformed all previous CNN models in efficiency and accuracy. A state-of-the-art deep neural network architecture, EfficientNet-B0, was used for this procedure. The model had 237 layers and 5,330,564 parameters, and 5,288,548 parameters were needed for gradient descent in training. The core structure of the network was the Mobile Inverted Bottleneck Convolution (MBConvBlock) module, which also uses the Squeeze-and-Excitation Network (SENet). When SENet was proposed, it achieved the highest accuracy on Imagenet data at that time. The deep CNN uses back propagation to train the model. All layers of the network were fine-tuned using AdamW (https://arxiV.org/pdf/1711.05101.pdf), a method for stochastic optimization with a global learning rate of 0.01 and decay to one-tenth every 30 epochs. To optimize images for EfficientNet-B0, they were resized to 512 × 512 pixels.The typical endoscopic pictures for *H. pylori* positive and negative are described in the Supplementary Material.

Statistical analysis

Estimation of sample size

The purpose was to verify the noninferiority of AI-assisted endoscopists compared with AI-unassisted endoscopists in the diagnosis of *H. pylori* infection. In the test data, AI diagnostic accuracy was 89.6%, and the accuracy of the endoscopists was 82.4% [14]. Based on a noninferiority margin of 10%, an α of 0.05, and a $1-\beta$ of 0.90, the estimated sample size was 84 per group. Noninferiority would be inferred if the 95% lower confidence boundary for the difference between the two groups in accuracy was more than 10%. The sample size was expanded using PASS version 11,with a total of 418 patients in the final two groups.

Analysis of outcomes

Numerical data were expressed as means ± SDs. The accuracy, specificity, sensitivity, PPV, and NPV from the different analyses were compared using the chi-square test or Fisher's exact test, as appropriate. All statistical calculations were performed using IBM SPSS version 23. A *p*-value below 0.05 was considered statistically significant. The trained CNN generated a continuous number for the probability of *H. pylori* infection (range: 0 to 1.0), and this was used as a continuous variable in the ROC analysis. The AUC values of the *H. pylori*-positive and *H. pylori*-negative groups were compared and plotted using R language version 4.01.

Results

Development data

From September 2020 to December 2020, 639 patients (260 *H. pylori*-positive, 379 *H. pylori*-negative; 289 males, 350 females) received EGDs in the endoscopy center of the Department of Gastroenterology, Daping Hospital of Chongqing (Table 1). For every patient, there was at least one image from each of five sites in the stomach (upper body, middle body, lower body, lesser curvature, and antrum). We used 7377 images from 639 patients (development data set) to construct the CNN.

Performance of the CNN

From January 2021 to March 2021, 201 different patients (66 *H. pylori*-positive, 135 *H. pylori*-negative; 90 males, 111 females) received EGDs at the same institution and these images were used for the test data set (Table 1). For the diagnosis of *H. pylori* infection, the CNN model had an accuracy of 89.6%, a sensitivity of 90.9%, and a specificity of 88.9%. The AUC for *H. pylori* positivity was 84.1% (95%CI: 73.0%, 95.2%) and the AUC for *H. pylori* negativity was 90.3% (95%CI: 82.2%, 98.4%; Fig. 2).

We also performed separate analyses of diagnosis based on the gastric body alone (upper body, middle body, lower body, and lesser curvature) and the antrum alone (Table 2). Analysis of the gastric body provided significantly greater sensitivity (86.2% *vs.* 19.8%) and accuracy (76.4% *vs.* 68.4%), but analysis of the antrum provided significantly greater specificity (93.3% *vs.* 71.7%); analysis of the antrum provided a significantly lower false-positive rate (6.7% *vs.* 28.3%), but analysis of the body provided a significantly lower false-negative rate (13.8% *vs.* 80.2%). Combining data from the body and antrum provided the best diagnostic performance.

Table 2 Sensitivity, specificity, accuracy, false-positive rate, and false-negative rate for diagnosis of *H. pylori* infection based on examination of the body and the antrum

Diagnostic parameter	Body[a]	Antrum	*P* value
Sensitivity	86.2%	19.8%	< 0.001
Specificity	71.7%	93.3%	< 0.001
Accuracy	76.4%	68.4%	< 0.001
False positive rate	28.3%	6.7%	< 0.001
False negative rate	13.8%	80.2%	< 0.001

[a] Stomach body consists of the upper body, middle body, lower body, and lesser curvature

Multicenter randomized controlled study

From April 2021 to July 2021, we enrolled 418 patients from three institutions (Department of Gastroenterology of Chongqing Daping Hospital, The 13th People's Hospital of Chongqing, and The Second People's Hospital of Jiulongpo District of Chongqing) for a multicenter randomized controlled trial of endoscopists. There were 209 patients in the AI-assisted group and 209 other patients in the AI-unassisted group (Table 3), and these two groups had no significant differences at baseline.

Comparisons of all 14 endoscopists in the AI-assisted and AI-unassisted groups indicated the AI-assisted group had significantly better accuracy (92.8% *vs.* 75.6%, $P<0.001$), sensitivity (91.8% *vs.* 78.6%, $P=0.032$), and specificity (93.4% *vs.* 74.5%, $P<0.001$; Table 4).

We then determined the accuracy of endoscopists in the AI-unassisted group who had different levels of experience, using two different metrics to define "junior" and "senior" status (Fig. 3). The accuracy of diagnosis was 85.5% (47/55) for senior endoscopists who were trained in early cancer detection, and was 72.1% (111/154)

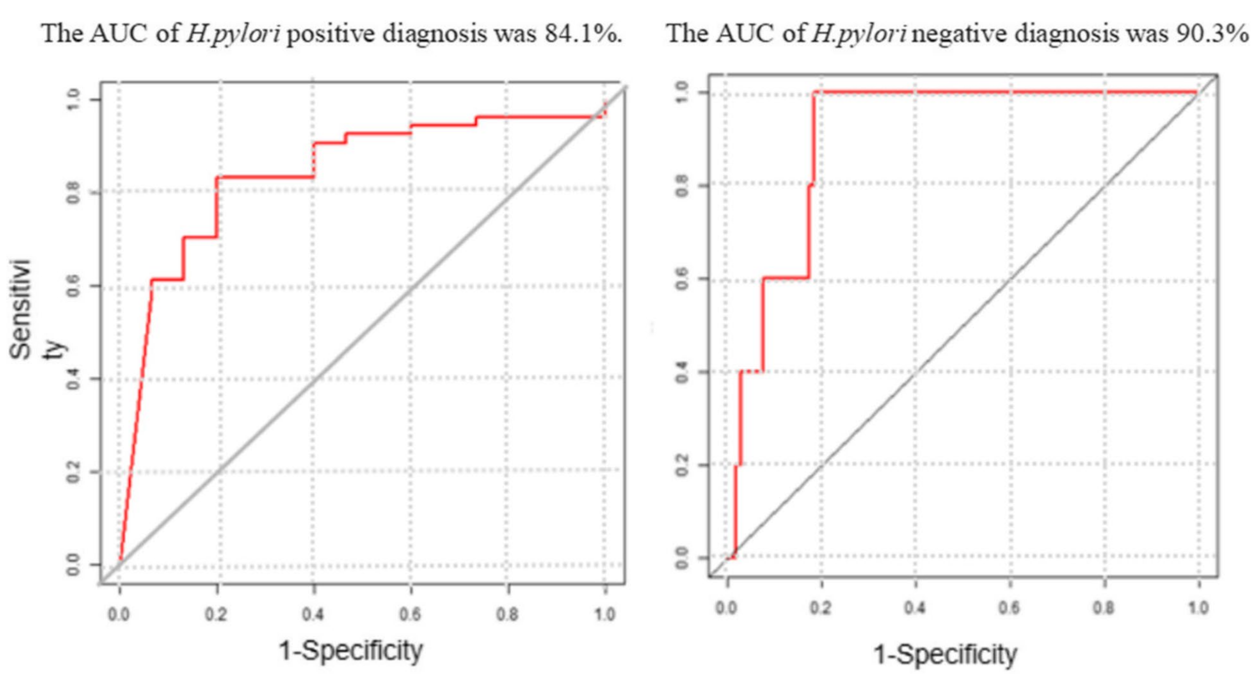

Fig. 2 Positive diagnosis (left) and negative diagnosis (right) of *H. pylori* infection using the CNN model

Table 3 Baseline demographic and clinical characteristics of patients examined in the multicenter randomized controlled study of endoscopists

Characteristic	AI-assisted	AI-unassisted	*p*-value
Sex			
Male, *n* (%)	92/209(44.0)	81/209(38.7)	0.275
Female, *n* (%)	117/209(56.0)	128/209(61.3)	
Mean age, years (SD)	46.3 ± 12.1	45.7 ± 12.4	0.652
Mean BMI (SD)	22.7 ± 3.4	23.2 ± 2.9	0.134
H. pylori status, *n* (%)			
Positive	73/209(34.9)	56/209(26.7)	0.072
Negative	136/209(65.0)	153/209(73.2)	
Cigarette smoking, *n* (%)	33/209(15.7)	28/209(13.3)	0.488
Alcohol drinking, *n* (%)	53/209(25.3)	47/209(22.4)	0.492
Family history of esophageal or gastric carcinoma, *n* (%)	12/209(5.7)	18/209(8.6)	0.256
Biopsy cases, *n* (%)	26/209(12.4)	31/209(14.8)	0.476
Atrophy cases, *n* (%)	61/209(29.1)	64/209(30.6)	0.749

for junior endoscopists who did not have this training ($P=0.047$). The accuracy of diagnosis was 77.3% (75/97) for senior endoscopists who had 5 years of experience or examination of more than 5000 cases, and was 73.2% (82/112) for junior endoscopists who did not have this experience ($P=0.494$).

Discussion

The development of endoscopic technology has made it possible to identify *H. pylori* using EGD. However, some current guidelines do not recommend endoscopy for the routine diagnosis of *H. pylori* infection because these methods require special equipment, endoscopists must receive relevant training, and the accuracy and specificity may differ among endoscopists [23]. Notably, developments in AI may resolve some of the problems related to the endoscopic diagnosis of *H. pylori* infection. AI methods can learn rules from sample data, and can recognize different types of data, such as text, images, and sounds. The image recognition used in the present study is based on deep learning and is a key technology for AI-assisted endoscopic diagnosis of *H.pylori* infection [12, 13]. Recent studies showed that AI endoscopic diagnosis of *H.pylori* infection is possible. In particular, a meta-analysis of AI-based diagnosis of *H.pylori* infection from endoscopy showed that the sensitivity was 0.87 (95%CI: 0.72, 0.94), the specificity was 0.86 (95%CI: 0.77, 0.92), and the AUC was 0.92 (95%CI: 0.90, 0.94) [15]. However, most studies of this topic were retrospective and diagnostic, and there have been no clinical prospective studies [15, 19–22]. In the present study, we combined a diagnostic study and multicenter randomized controlled study of endoscopists. Our results confirmed the clinical value of AI-assisted endoscopic diagnosis of *H. pylori* infection.

In the present study, the endoscopic diagnostic model of *H. pylori* infection was based on deep learning and was constructed using endoscopic images of five stomach regions. In the test data, AI diagnosis that was based on images of the gastric body (upper body, middle body, lower body, and lesser curvature) provided better accuracy, sensitivity, and PPV than AI diagnosis based on images of the antrum; however, images of the antrum provided better specificity and NPV than images of the body. The combined use of body and antrum images had an accuracy of 89.6%, a sensitivity of 90.9%, and a specificity of 88.9%. A 2020 Japanese study found that diagnostic accuracy was greater when using the lesser curvature of the middle-upper body than the fornix and the greater curvature of the middle-upper body [17, 18]. A 2019 study in China found that the sensitivity, specificity, and accuracy from using multiple gastric images (average

Table 4 Sensitivity, specificity, accuracy, positive predictive value (PPV), and negative predictive value (NPV) of endoscopists with and without AI assistance in the diagnosis of *H. pylori* infection

	AI-assisted		Total	AI-unassisted		Total	
Diagnostic parameter	Untrained[a] ($n=11$)	Trained[b] ($n=3$)	14	Untrained[a] ($n=11$)	Trained[b] ($n=3$)	14	*P* value[***]
Sensitivity	91.6%(44/48)	92.0%(23/25)	91.8%(67/73)	76.3%(29/38)	83.3%(15/18)	78.6%(44/56)	**0.032**
Specificity	91.9%(103/112)	100%(24/24)	93.4%(127/136)	70.6%(82/116)	86.4%(32/37)	74.5%(114/153)	**< 0.001**
Accuracy	91.9%(147/160)	95.9%(47/49)	92.8%(194/209)	72.1%(111/154)	85.5%(47/55)	75.6%(158/209)	**< 0.001**
PPV	83.0%(44/53)	100%(23/23)	88.2%(67/76)	46.0%(29/63)	75.0%(15/20)	53.0%(44/83)	**< 0.001**
NPV	96.2%(103/107)	92.3%(24/26)	95.5%(127/133)	90.1%(82/91)	91.4%(32/35)	90.5%(114/126)	**0.113**

[a] Not trained for early cancer screening

[b] Trained for early cancer screening

[***] Comparison of all AI-assisted *vs.* all AI-unassisted

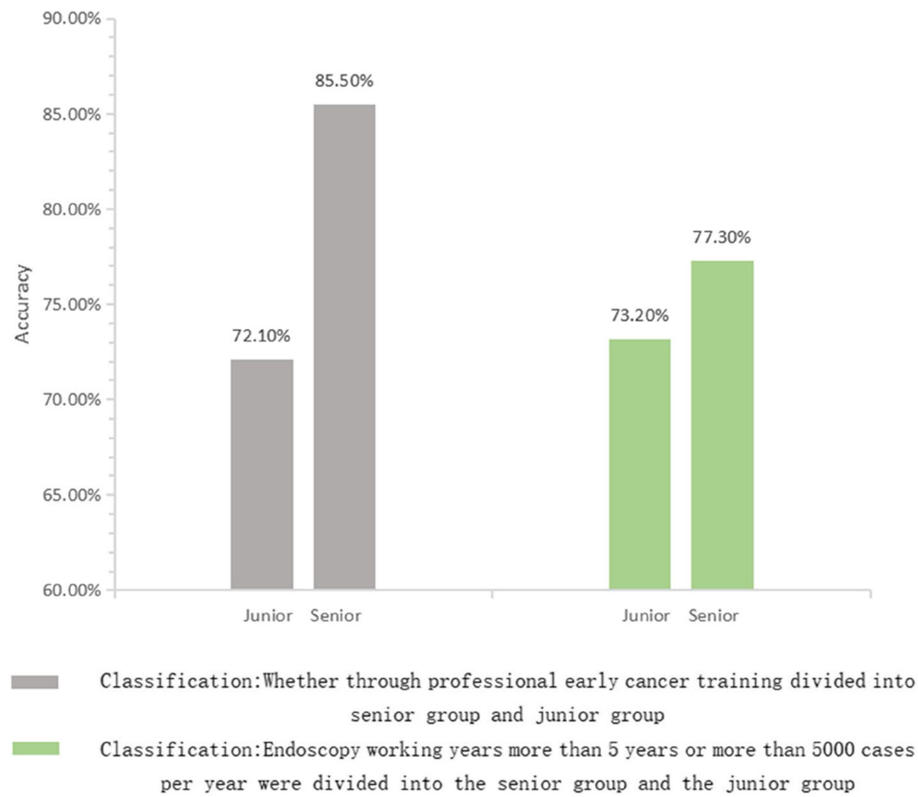

Fig. 3 Accuracy of endoscopists diagnosis of *H. pylori* infection when using different criteria to define "junior" and "senior" endoscopists

8.3 ± 3.3 per patient) for AI diagnosis of *H. pylori* infection were greater than those from using a single image [18]. This led us to use an AI-assisted endoscopy diagnostic system for *H. pylori* infection that is based on images of multiple sites for model establishment.

Most previous endoscopy studies have used WLI, and only a few have used newer endoscopic techniques. A 2018 Japanese study compared the effectiveness of using AI with WLI, Blue Laser Imaging (BLI), and Linked Color Imaging (LCI) for the diagnosis of *H. pylori* infection, and found that the AUC was higher for BLI (0.96) and LCI (0.95) than for WLI (0.66, both $P < 0.01$) [16]. Although these new techniques seem to have advantages over WLI in the diagnosis of *H. pylori* infection, they have not been widely adopted because they require additional expensive equipment and professionally trained endoscopists. Therefore, an *H. pylori* diagnostic system developed for WLI is more suitable for widespread acceptance, especially in regions where resources are limited. In addition, some of the limitations of WLI can be overcome by use of AI learning from images at multiple sites in the stomach.

The present study was the first to examine an AI system for the endoscopic diagnosis of *H. pylori* infection in a multicenter randomized controlled clinical trial of endoscopists. We found that endoscopists with AI assistance had significantly better accuracy, sensitivity, and specificity than unassisted endoscopists (all $P < 0.05$). This indicates that our AI endoscopic system for the diagnosis of *H. pylori* infection significantly improved the diagnostic ability of endoscopists when using gastroscopy.

Our analysis of endoscopists also found the number of working years and the number of procedures that were performed had no significant effect on their diagnostic performance. However, endoscopists who were trained in early cancer detection had greater accuracy than endoscopists without this training (85.5% [47/55] *vs.* 72.1% [111/154], $P = 0.047$). This may be because during training in the early stages of cancer, endoscopic identification of the *H.pylori* status of the gastric mucosa is important in determining the mucosal background of gastric cancer, and thus these professional trained endoscotists are familiar with the manifestations of *H.pylori* infection of the gastric mucosa. Endoscopists not trained in early cancer detection may have lower accuracy, even though they are skilled in endoscopy procedures. This further emphasizes the clinical value of the using an AI system for the endoscopic diagnosis of *H. pylori* infection.

The UBT is the most common method for the clinical detection of *H. pylori* infection. However, when the results of this test are close to the test threshold, diagnosis can be difficult. The present study showed that an AI endoscopy system can assist doctors whose UBT results are near the threshold. In particular, among the 418 patients, 6 patients had UBT results near the critical value, and 5 of them were accurately assessed by the AI system, corresponding to an accuracy of 83.3% and a sensitivity of 100% (data was not included in study as it only resulted from 6 patients). However, this result is from only 6 patients, so a larger sample size is needed for further study of this topic.

The status of stomach infection by *H. pylori* may be classified as "infected", "uninfected", or "eradicated", with the latter two states considered negative. We found it was difficult to identify the eradicated state using endoscopy, and identification of this state was also affected by the time since eradication. In particular, among the 418 patients, 369 were not treated for *H. pylori* infection, and the accuracy of diagnosis in these patients was 85.9%, greater than the accuracy in patients treated for *H. pylori* infection (71.4%, data not shown). This result indicates that prior treatment for *H. pylori* infection adversely affected diagnosis by doctors alone and by doctors using AI. This may be because of decreased gastric mucosal inflammation and atypical gastric mucosal appearance after the use of antibiotics and PPIs. We also found better accuracy (81.0% *vs.* 65.4%), specificity (83.3% *vs.* 68.1%), and sensitivity (66.7% *vs.* 50.0%) in patients who received treatment more than two years ago rather than less than 2 years ago (data not shown). We speculate that as the time after treatment increases, there were greater declines in inflammation of the gastric mucosa,so endoscopic gastric mucosa becomes more typical [24–27]. For example, the diffuse redness from the gastric body to the gastric fundus disappears, and some of these individuals even show RAC again, so the accuracy of endoscopic diagnosis improved accordingly.

As the most convenient non-invasive examination method, UBT is still the first choice for *H. pylori* detection in patients who do not undergo gastroscopy. In China, the cost of an endoscopic examination is relatively low,about 200 RMB,and it is widely accepted by patients. Because of china has a high rate of infection by *H. pylori* and a high incidence of gastric cancer,gastroscopy is recommended as one of the routine physical examinations for people over 40 years old,if two tests can be combined and completed simultaneously using gastroscopy, *H.pylori* infection can be treated in a timely manner, precancerous lesions of gastric cancer and early gastric cancer can be detected, and medical costs can be reduced.

There are some limitations in this study. First, our development and test data sets were all from a single center. Second, our sample size was rather small, and a larger sample should be used in subsequent studies. Finally, this study did not use the classification system of "positive", "negative", and "negative after eradication", a topic that is also worthy of further study.

In conclusion, we established an AI-assisted endoscopy system for the diagnosis of *H. pylori* infection that was based on AI learning of images from five sites in the stomach, and applied this system to the first clinical trial of this topic in China. The results of this multicenter randomized controlled trial verified that the AI system described here improved the ability of endoscopists to diagnose *H. pylori* infection using gastroscopy. Because this system provided greater improvements to endoscopists who were not trained in early cancer detection, we believe it may be beneficial for geographic regions that have limited resources, high incidences of gastric cancer, and common use and acceptance of gastroscopy.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12876-024-03389-3.

Supplementary Material 1.

Authors' contributions

All authors contributed to the study conception and design. The first three authors contributed equally to this work. Material preparation, data collection, and data analysis were performed by Pei-Ying Zou, Jian-Ru Zhu,Zhe Zhao,Hao Mei,Jing-Tao Zhao,Wen-Jing Sun,Guo-Hua Wang,Dong-Feng Chen and Li-Lin Fan.The first draft of the manuscript was written by Pei-Ying Zou. Chun-Hui Lan critically reviewed the manuscript. All the authors commented on previous versions of the manuscript and approved the final manuscript.

Funding

This study was supported by grants from the Science and Technology Innovation Enhancement Project of Army Medical University (No. 20I9CXLCB003) and the National Science Foundation of China (No. 82072253).

Availability of data and materials

All data generated or analysed during this study are included in this published article.

Declarations

Ethics approval and consent to participate

The experimental protocol was established, according to the ethical guidelines of the Helsinki Declaration and was approved by the Human Ethics Committee of Chinese Clinical Trial Registry (www.chictr.org.cn ;registration number: ChiCTR2000037801). This study was approved by the Ethics Committee of Daping Hospital (10/07/2020) (No.89,2020).Informed consent was obtained from all study participants.

Consent for pulication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details
[1]Department of Gastroenterology, Chongqing Key Laboratory of Digestive Malignancies, Daping Hospital, Army Medical University, Third Military Medical University), 10 Changjiang Branch Road, Chongqing 400000, China. [2]Chongqing 13, People's Hospital, Chongqing, China. [3]ChongqingSkyforbioCo, Ltd.MR, Chongqing, China. [4]Chongqing Jiulongpo District Second People's Hospital, Chongqing, China.

Received: 13 December 2023 Accepted: 27 August 2024
Published online: 30 September 2024

References

1. Hooi JKY, et al. Global Prevalence of Helicobacter pylori Infection: Systematic Review and Meta-Analysis. Gastroenterology. 2017;153(2):420–9.
2. Savoldi A, et al. Prevalence of Antibiotic Resistance in Helicobacter pylori: A Systematic Review and Meta-analysis in World Health Organization Regions. Gastroenterology. 2018;155(5):1372-1382.e17.
3. Ndip RN, et al. Helicobacter pylori isolates recovered from gastric biopsies of patients with gastro-duodenal pathologies in Cameroon: current status of antibiogram. Tropical Med Int Health. 2008;13(6):848–54.
4. Huh CW, Kim BW. Diagnosis of Helicobacter pylori Infection. Korean J Gastroenterol. 2018;72(5):229–36.
5. Wang YK, et al. Diagnosis of Helicobacter pylori infection: Current options and developments. World J Gastroenterol. 2015;21(40):11221–35.
6. Makristathis A, et al. Review: Diagnosis of Helicobacter pylori infection. Helicobacter. 2019;24(Suppl 1):e12641.
7. Huangwei C, et al. The application value of Kyoto gastritis classification in the direct judgment of Helicobacter pylori infection under white light gastroscopy. New Med. 2019;50(6):457–62.
8. Motoji K, et al. Classification of gastritis by Kyoto. Stomach and Intestine. 2019;54(5):616–9.
9. Kato T, Yagi N, Kamada T, Shimbo T, Watanabe H, Ida K. Diagnosis of Helicobacter pylori infection in gastric mucosa by endoscopic features: a multicenter prospective. Dig Endosc. 2013;25(5):508–18.
10. Cho JH, Chang YW, Jang JY, Shim JJ, Lee CK, Dong SH, Kim HJ, Kim BH, Lee TH, Cho JY. Close observation of gastric mucosal pattern by standard endoscopy can predict Helicobacter pylori infection status. J Gastroenterol Hepatol. 2013;28(2):279–84.
11. Watanabe K, Nagata N, Nakashima R, Furuhata E, Shimbo T, Kobayakawa M, Sakurai T, Imbe K, Niikura R, Yokoi CJ, Uemura N. Predictive findings for Helicobacter pylori-uninfected,-infected and -eradicated gastric mucosa: validation study. World J Gastroenterol. 2013;19(27):4374–9.
12. Esteva A, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542(7639):115–8.
13. Ribeiro E, Uhl A, Wimmer G, Häfner M. Exploring deep learning and transfer learning for colonic polyp classification. Comput Math Methods Med. 2016;2016:6584725.
14. Shichijo S, et al. Application of Convolutional Neural Networks in the Diagnosis of Helicobacter pylori Infection Based on Endoscopic Images. EBioMedicine. 2017;25:106–11.
15. Bang CS, Lee JJ, Baik GH. Artificial Intelligence for the Prediction of Helicobacter Pylori Infection in Endoscopic Images: Systematic Review and Meta-Analysis Of Diagnostic Test Accuracy. J Med Internet Res. 2020;22(9):e21983.
16. Nakashima H, et al. Artificial intelligence diagnosis of Helicobacter pylori infection using blue laser imaging-bright and linked color imaging: a single-center prospective study. Ann Gastroenterol. 2018;31(4):462–8.
17. Yasuda T, et al. Potential of automatic diagnosis system with linked color imaging for diagnosis of Helicobacter pylori infection. Dig Endosc. 2020;32(3):373–81.
18. Zheng W, et al. High Accuracy of Convolutional Neural Network for Evaluation of Helicobacter pylori Infection Based on Endoscopic Images: Preliminary Experience. Clin Transl Gastroenterol. 2019;10(12):e00109.
19. Shichijo S, Endo Y, Aoyama K, Takeuchi Y, Ozawa T, Takiyama H, et al. Application of convolutional neural networks for evaluating Helicobacter pylori infection status on the basis of endoscopic images. Scand J Gastroenterol. 2019;54(2):158–63.
20. Itoh T, Kawahira H, Nakashima H, Yata N. Deep learning analyzes Helicobacter pylori infection by upper gastrointestinal endoscopy images. Endosc Int Open 2018 Feb;6(2):E139-E144 [FREE Full text] https://doi.org/10.1055/s-0043-120830] [Medline: 29399610].
21. Huang C, Chung P, Sheu B, Kuo H, Popper M. Helicobacter pylori-related gastric histology classification using support-vector-machine-based feature selection. IEEE Trans Inf Technol Biomed. 2008;12(4):523–31.
22. Huang C, Sheu B, Chung P, Yang H. Computerized diagnosis of Helicobacter pylori infection and associated gastric inflammation from endoscopic images by refined feature selection using a neural network. Endoscopy. 2004Jul;36(7):601–8.
23. Wenzhong L, et al. The fifth National Consensus report on the management of Helicobacter pylori infection. Gastroenterology. 2017;22(06):346–60.
24. Zhang Y, Jing D, Qiu Y. Effects of Helicobacter pylori eradication on gastric mucosal precancerous lesions: a meta-analysis. J Clin Gastroenterol. 2009;21(05):268–72.
25. Sung JJ, Lin SR, Ching JY, et al. Atrophy and intestinal metaplasia one year after cure of H. pylori infection: a prospective, randomized study. Gastroenterology. 2000;119:7–14.
26. Bravo JC, et al. Chemoprevention of gastric dysplasia: randomized trial of antioxidants supplements and anti- Helicobacter pylori therapy. J Natl Cancer Inst. 2000;92:1881–8.
27. Khan MY, et al. Effectiveness of Helicobacter pylori eradication in preventing metachronous gastric cancer and preneoplastic lesions. A systematic review and meta-analysis. Eur J Gastroenterol Hepatol. 2020;32(6):686–94.

Publisher's Note