

# High Accuracy of Convolutional Neural Network for Evaluation of *Helicobacter pylori* Infection Based on Endoscopic Images: Preliminary Experience

Wenfang Zheng, BM<sup>1,2</sup>, Xu Zhang, BE<sup>3</sup>, John J. Kim, MD, MS<sup>1,4</sup>, Xinjian Zhu, MD<sup>5</sup>, Guoliang Ye, MD<sup>6</sup>, Bin Ye, MD<sup>7</sup>, Jianping Wang, MD<sup>8</sup>, Songlin Luo, MD<sup>9</sup>, Jingjing Li, MM<sup>10</sup>, Tao Yu, BE<sup>3</sup>, Jiquan Liu, PhD<sup>3</sup>, Weiling Hu, MD, PhD<sup>1,2</sup> and Jianmin Si, MD, PhD<sup>1,2</sup>

- OBJECTIVES:** Application of artificial intelligence in gastrointestinal endoscopy is increasing. The aim of the study was to examine the accuracy of convolutional neural network (CNN) using endoscopic images for evaluating *Helicobacter pylori* (*H. pylori*) infection.
- METHODS:** Patients who received upper endoscopy and gastric biopsies at Sir Run Run Shaw Hospital (January 2015–June 2015) were retrospectively searched. A novel Computer-Aided Decision Support System that incorporates CNN model (ResNet-50) based on endoscopic gastric images was developed to evaluate for *H. pylori* infection. Diagnostic accuracy was evaluated in an independent validation cohort. *H. pylori* infection was defined by the presence of *H. pylori* on immunohistochemistry testing on gastric biopsies and/or a positive 13C-urea breath test.
- RESULTS:** Of 1,959 patients, 1,507 (77%) including 847 (56%) with *H. pylori* infection (11,729 gastric images) were assigned to the derivation cohort, and 452 (23%) including 310 (69%) with *H. pylori* infection (3,755 images) were assigned to the validation cohort. The area under the curve for a single gastric image was 0.93 (95% confidence interval [CI] 0.92–0.94) with sensitivity, specificity, and accuracy of 81.4% (95% CI 79.8%–82.9%), 90.1% (95% CI 88.4%–91.7%), and 84.5% (95% CI 83.3%–85.7%), respectively, using an optimal cutoff value of 0.3. Area under the curve for multiple gastric images (8.3 ± 3.3) per patient was 0.97 (95% CI 0.96–0.99) with sensitivity, specificity, and accuracy of 91.6% (95% CI 88.0%–94.4%), 98.6% (95% CI 95.0%–99.8%), and 93.8% (95% CI 91.2%–95.8%), respectively, using an optimal cutoff value of 0.4.
- DISCUSSION:** In this pilot study, CNN using multiple archived gastric images achieved high diagnostic accuracy for the evaluation of *H. pylori* infection.

Clinical and Translational Gastroenterology 2019;10:e00109. <https://doi.org/10.14309/ctg.000000000000109>

## INTRODUCTION

*Helicobacter pylori* (*H. pylori*) infects the epithelial lining of the stomach and is associated with functional dyspepsia, peptic ulcers, and gastric cancer (1). Endoscopy is frequently performed for the evaluation of *H. pylori*-associated diseases, and certain endoscopic features are associated with *H. pylori* infection (2). However, evaluation of *H. pylori* at the time of endoscopy requires gastric biopsies because endoscopic impression alone is inaccurate (3).

Emerging studies have highlighted the application of artificial intelligence in gastrointestinal endoscopy (4). Convolutional neural network (CNN), architecture for deep learning in medical image analysis, has been evaluated in gastrointestinal disease (5–7). Discriminating endoscopic features can be extracted by CNN at multiple levels of abstraction in a large data set to derive a model to provide a probability for the presence of pathology. Given remarkable visual recognition capability, we hypothesize that CNN

<sup>1</sup>Department of Gastroenterology, Sir Run Run Shaw Hospital, Medical School, Zhejiang University, Hangzhou, China; <sup>2</sup>Institute of Gastroenterology, Zhejiang University (IGZJU), Hangzhou, China; <sup>3</sup>Key Laboratory for Biomedical Engineering of Ministry of Education, College of Biomedical Engineering & Instrument Science, Zhejiang University, Hangzhou, China; <sup>4</sup>Division of Gastroenterology and Hepatology, Loma Linda University Health, Loma Linda, California, USA; <sup>5</sup>Department of Gastroenterology, Shaoxing Shangyu People's Hospital and Shangyu Hospital of the Second Affiliated Hospital, Medical School, Zhejiang University, Shaoxing, China; <sup>6</sup>Department of Gastroenterology, Affiliated Hospital, Medical School, Ningbo University, Ningbo, China; <sup>7</sup>Department of Gastroenterology, Fifth Affiliated Hospital of Wenzhou Medical University and Lishui Municipal Central Hospital, Lishui, China; <sup>8</sup>Department of Gastroenterology, Deqing People's Hospital, Huzhou, China; <sup>9</sup>Department of Gastroenterology, Shangyu Hospital of Traditional Chinese Medicine, Shaoxing, China; <sup>10</sup>Department of Gastroenterology, First People's Hospital of Huzhou, Huzhou, China. **Correspondence:** Weiling Hu, PhD. E-mail: huweiling@zju.edu.cn. Jiquan Liu, PhD. E-mail: liujq@zju.edu.cn. **Received June 17, 2019; accepted October 21, 2019; published online December 11, 2019**

technology can accurately evaluate for *H. pylori* infection during conventional endoscopy without the need for biopsies. We have developed Computer-Aided Decision Support System that uses CNN to evaluate for *H. pylori* infection based on endoscopic images. The aim of the study was to evaluate the accuracy of CNN to evaluate for *H. pylori* infection based on archived endoscopic images.

## METHODS

### Patient population

Patients receiving endoscopy with gastric biopsies at Sir Run Run Shaw Hospital (Hangzhou, China) from January 2015 to June 2015 were retrospectively searched. Patients with a history of gastric cancer, peptic ulcers, or submucosal tumor, as well as, having endoscopic findings of ulcer, mass, or strictures, were excluded. Furthermore, patients who had antibiotics within a month or proton pump inhibitor within 2 weeks of endoscopy were excluded by reviewing medical records. Immunohistochemistry testing was performed in all gastric biopsy specimens to evaluate for *H. pylori* infection. If a patient had no evidence of *H. pylori* infection on gastric biopsies, only those who had *H. pylori* breath test performed within a month before or after the endoscopy in the absence of documented eradication treatment were included. The endoscopic images of the study population derived from January 2015 to May 2015 were assigned to the derivation group for machine learning using computer-aided decision support system. The remaining study population who received endoscopy in June 2015 was assigned to the validation group to evaluate the accuracy of computer-aided decision support system-derived model for *H. pylori* evaluation. The study was approved by the Ethics Committee of Sir Run Run Shaw Hospital, College of Medicine, Zhejiang University (20190122-8), before initiating the study.

### Upper endoscopy examination

Upper endoscopy was performed using a standard endoscope (GIF-Q260J; Olympus, Tokyo, Japan). Gastric images captured during high-definition, white-light examination of the antrum, angularis (retroflex), body (forward and retroflex), and fundus (retroflex) were used for both the derivation and validation sets. Gastric biopsies were obtained in the antrum and/or body per discretion of the endoscopist.

### Data

Archived gastric images obtained during standard white-light examination from the endoscopic database were extracted. Two endoscopists independently screened and excluded images that were suboptimal in quality (i.e., blurred images, excessive mucus, food residue, bleeding, and/or insufficient air insufflation). Selected images were randomly rotated between 0° and 359° for data augmentation to improve the accuracy of the model trained by CNN (8).

### Training algorithm

The Computer-Aided Decision Support System (College of Biomedical Engineering & Instrument Science, Zhejiang University, Hangzhou, China) that uses ResNet-50 (Microsoft), a state-of-the-art CNN consisting of 50 layers, was developed. PyTorch (Facebook) as a deep learning framework known for flexibility and conduciveness to train CNN was used. Stochastic gradient descent algorithm with back propagation was used to update the weights of the model. The momentum was set at 0.9

and weight decay at 0.0001. The initial learning rate was set at 0.0005, which was reduced 10 times after every 8 epochs for a total of 20 epochs. Cross entropy was applied as a loss function. Each image was resized to 224 × 224 pixels to fit ResNet requirement. A pretrained model that learns natural-image features from ImageNet was used to accelerate model convergence and alleviate model overfitting problems caused by insufficient derivation set images (9). Three deep learning models were created using the data from the derivation cohort: (i) single gastric image for all gastric images; (ii) single gastric image by different gastric locations (fundus, corpus, angularis, and antrum); and (iii) multiple gastric images for the same patient.

### Evaluation algorithm

After the deep learning model was constructed, a validation cohort was used to evaluate the diagnostic accuracy CNN for the evaluation of *H. pylori* infection. The Convolutional Neural Network outputs a continuous number between 0 and 1 for the probability of *H. pylori* infection.

### Definition and study endpoints

*H. pylori* infection was defined by the presence of *H. pylori* on immunohistochemistry testing from gastric biopsies and/or a positive C13 breath test. Study endpoints were sensitivity, specificity, and accuracy of the trained CNN by using single and multiple endoscopic images. Furthermore, subgroup analyses evaluating sensitivity, specificity, and accuracy by anatomical locations (antrum, angularis, body, and fundus) and by the mode of gastric biopsies (antrum biopsy alone, antrum, and body biopsies) were calculated.

### Statistics

Demographic data were expressed as mean with SD or median with ranges as appropriate. Sensitivity, specificity, and accuracy with 95% confidence interval [CI] were calculated by using SDs. Receiver operating characteristic (ROC) curves were plotted, and the area under the curve (AUC) with 95% CI was calculated. Optimal cutoff values to obtain the highest AUC were calculated using the Youden index. Two AUCs of models were compared using the DeLong test. SPSS 23.0 and R was used for all statistical analyses. Two-side *P* value of < 0.05 was considered statistically significant.

## RESULTS

### Patient characteristics

During the study period, 9,066 patients received upper endoscopy with gastric biopsies. Nine hundred forty-five patients were excluded including those with gastric cancer in 41, gastric ulcers in 902, and submucosal tumor in 2. Furthermore, after excluding patients without evidence of *H. pylori* on gastric biopsy and did not receive a urea breath test (N = 6,162), 1,959 were eligible per the study criteria. The mean age of 1,959 patients was 48.5 ± 12.5, 956 (49%) were men, and 1,852 (95%) received outpatient procedures (Table 1). In addition to gastric biopsies with immunostain, 849 (43%) had an additional *H. pylori* breath test. A total of 1,157 (59%) patients had *H. pylori* infection documented on gastric biopsies and/or *H. pylori* breath test. Overall, 15,484 archived gastric images were available for analysis after excluding the suboptimal images (N = 5,064).

Using June 1, 2015, as the demarcation point, 1,507 (77%) of 1,959 patients (11,729 gastric images) were assigned to the derivation cohort, and 452 (23%) patients (3,755 gastric images including antrum in 869 [23%], angularis in 719 [19%], body in

**Table 1. Baseline characteristics**

	All N = 1,959	Derivation cohort N = 1,507	Validation cohort N = 452
Age, yrs (SD)	48.5 ± 12.5	48.3 ± 12.2	48.6 ± 12.9
Male (%)	956 (48.8)	736 (48.8)	220 (48.7)
Indication for gastroscopy			
Non-ulcer dyspepsia	686	528	158
GERD symptoms	230	183	47
Screening/surveillance	713	552	161
Others	330	244	86
Original gastric images	15,484	11,729	3,755
Images after data augmentation	79,901	76,146	3,755
Images per patient (SD)	8.1 ± 3.3	7.8 ± 3.3	8.3 ± 3.3
<i>H. pylori</i> infection (%)	1,157 (59.1)	847 (56.2)	310 (68.6)
GERD, gastroesophageal reflux disease.			

1,281 [34%], and fundus in 886 [24%]) were assigned to the validation cohort.

#### Performance of CNN for a single gastric image

Applying the CNN model in the validation set, the AUC for a single gastric image was 0.93 (95% CI 0.92–0.94). Using an optimal cutoff value of 0.28, sensitivity, specificity, and accuracy were 81.4% (95% CI 79.8%–82.9%), 90.1% (95% CI 88.4%–91.7%), and 84.5% (95% CI 83.3%–85.7%), respectively (Table 2).

When evaluating a single gastric image at different anatomical locations, the AUCs for the ROC ranged from 0.82 (95% CI 0.79–0.84) in the fundus to 0.94 (95% CI 0.92–0.95) in the body

**Table 2. Diagnostic characteristics of a single gastric image at difference cutoff values**

Cutoff value (probability score)	Accuracy (%)	Sensitivity (%)	Specificity (%)
0.1	85.0	87.0	81.2
0.2	85.1	83.6	87.6
0.28 <sup>a</sup>	84.5	81.4	90.1
0.3	84.2	80.7	90.4
0.4	83.4	78.4	92.5
0.5	82.6	76.2	94.0
0.6	81.5	73.4	96.0
0.7	80.0	70.5	97.1
0.8	77.8	66.3	98.1
0.9	73.9	60.0	98.8

<sup>a</sup>Optimal cutoff value was calculated using the Youden index.

(Figure 1 and Table 3). The CNN model using a single body image had the highest AUC ( $P < 0.01$ ) compared with the antrum or fundus (Table 3).

#### Performance of CNN for multiple gastric images

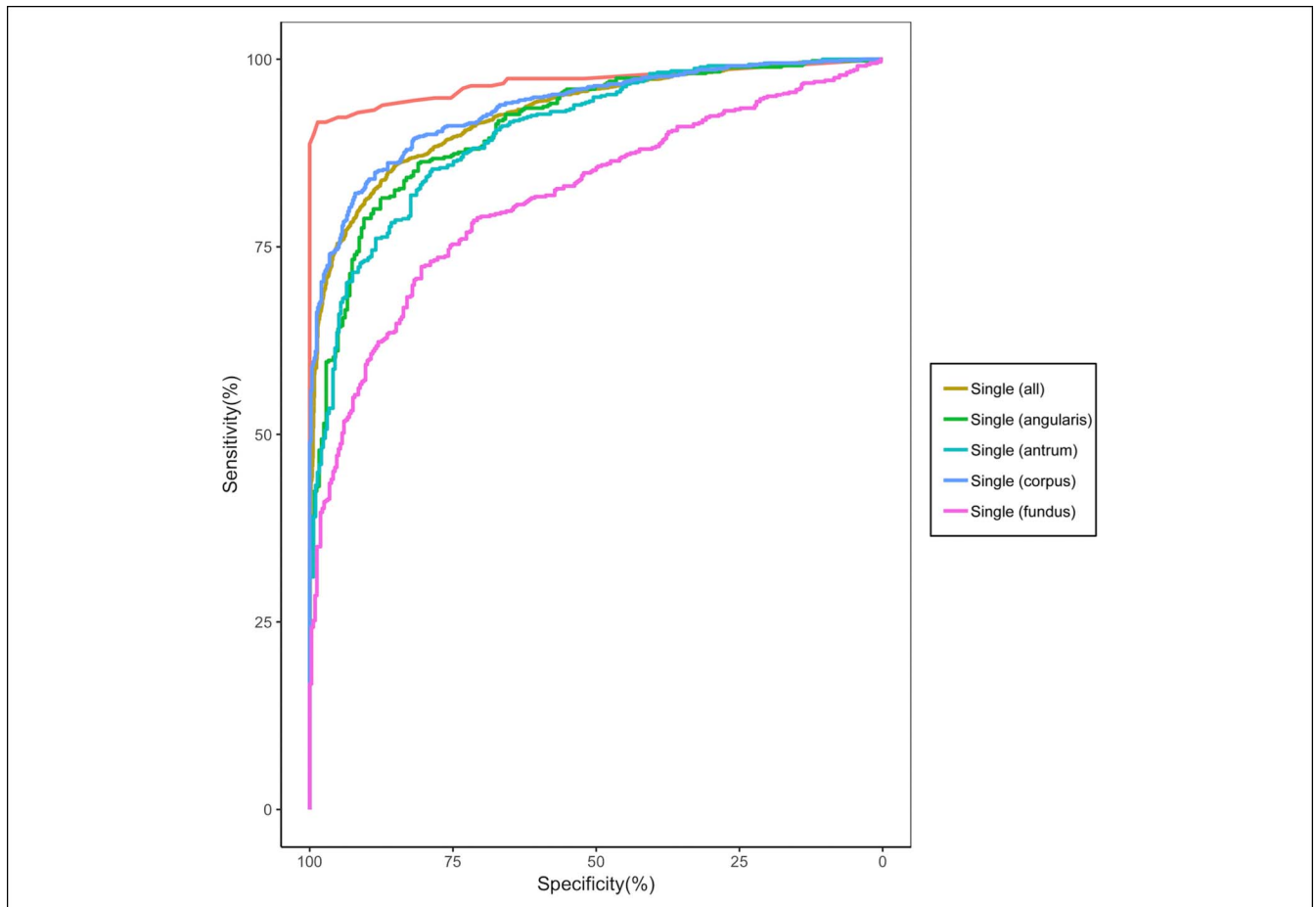
In the validation cohort, a mean of  $8.3 \pm 3.3$  gastric images was analyzed for each patient. Applying the CNN model for multiple gastric images per patient, the AUC for the ROC was 0.97 (95% CI 0.96–0.99) (Table 4). Using an optimal cutoff value of 0.4, the sensitivity, specificity, and accuracy were 91.6% (95% CI 88.0%–94.4%), 98.6% (95% CI 95.0%–99.8%) and 93.8% (95% CI 91.2%–95.8%), respectively. The CNN model using multiple gastric images had a higher AUC compared with a single gastric image ( $P < 0.001$ ) or body gastric image ( $P < 0.001$ ). In the subgroup analysis, sensitivity, specificity, and accuracy of CNN using multiple images for 398 (88%) patients receiving antrum biopsies alone were 91.7% (95% CI 87.9%–94.6%), 98.2% (95% CI 93.5%–99.8%), and 93.5% (95% CI 90.6%–95.7%), while 90.5% (95% CI 89.4%–91.7%), 100% (95% CI 89.4%–100%), and 96.3% (95% CI 87.3%–99.5%) for 54 (12%) patients who received antrum and body biopsies. No differences in sensitivity (mean difference = 1.2%, 95% CI –11.1% to 13.5%), specificity (mean difference = –1.8%, 95% CI –6.4% to 2.8%), or accuracy (mean difference = –2.8%, 95% CI –9.7% to 4.0%) were observed among patients receiving antrum biopsies alone compared with antrum and body biopsies.

#### DISCUSSION

In a single-center study of patients receiving endoscopy with gastric biopsies, CNN based on a single gastric image demonstrated modest accuracy of 81.4% (95% CI 79.8%–82.9%) and an AUC of 0.93 (95% CI 0.92–0.94) for evaluation of *H. pylori* infection. When comparing gastric locations, the body had the highest accuracy of 85.6% (83.6%–87.5%) and an AUC of 0.94 (95% CI 0.92–0.95) for detecting *H. pylori* infection. CNN model using multiple gastric images ( $8.3 \pm 3.3$ ) per patient demonstrated high AUC of 0.97 (95% CI 0.96–0.99) with sensitivity, specificity, and accuracy, exceeding 90% for evaluation of *H. pylori* infection.

*H. pylori* testing is recommended in patients with functional dyspepsia, atrophic gastritis, peptic ulcer disease, and gastric mucosa-associated lymphoid tissue lymphoma (10). Endoscopy is the primary diagnostic test for the evaluation of *H. pylori*-associated disease, and gastric biopsies are routinely obtained to evaluate for the presence of *H. pylori* infection.

The remarkable progress in artificial intelligence and visual recognition technology have led to increased application in a wide array of disciplines. CNN is an artificial intelligence network developed for detection, segmentation, and recognition of image patterns. For example, CNN system has been used for facial recognition technology for Facebook and Google images tool (11). After recognizing visual patterns from raw image pixel, a mathematical operation called convolution is used to classify the images. The application of CNN has been rapidly propagating in clinical medicine for computer-aided diagnosis, radiomics, and medical imaging analysis (12). In gastrointestinal endoscopy, CNN has been evaluated for detection of colon polyps during colonoscopy (5), evaluation of bleeding source on video capsule endoscopy (6), and assessment of depth of invasion in gastric cancer during upper endoscopy (7). With sufficient derivation image data, CNN capable of image recognition can facilitate endoscopists with diagnostic and therapeutic interventions.



**Figure 1.** Receiver operating curves for the CNN model using all single gastric images, single gastric image categorized by locations, and multiple gastric images.

Given well-described endoscopic features associated with the presence (e.g., gastric mucosal erythema, edema, and antral nodularity) and absence (i.e., regular arrangement of collecting venules) of *H. pylori* infection (2), we hypothesized that CNN can accurately evaluate the presence of *H. pylori* infection without the need for gastric biopsies.

In our study, the sensitivity, specificity, and accuracy of CNN model evaluating a single gastric image were 81.4% (95% CI 79.8%–82.9%), 90.1% (95% CI 88.4%–91.7%), and 84.5% (95% CI 83.3%–85.7%), respectively. In the subgroup analysis, a single

gastric body image had the highest AUC compared with a single gastric image of antrum or fundus. The diagnostic accuracy of a single gastric image in our study is within the range reported in previous studies. For example, a study of CNN model using 149 single gastric images demonstrated sensitivity and specificity of 87% in a validation cohort of 30 patients (13). In validation cohort of 30 patients, sensitivity and specificity were 86.7%. Furthermore, a 22-layer CNN pretrained on 32,208 gastric images categorized by 8 anatomical locations demonstrated sensitivity of 88.9%, specificity of 87.4%, and accuracy of 87.7%

**Table 3.** Diagnostic characteristic of single gastric image by location

	All, N = 3,755 (95% CI)	Antrum, N = 869 (95% CI)	Angularis, N = 719 (95% CI)	Corpus, N = 1,281 (95% CI)	Fundus, N = 886 (95% CI)
Accuracy (%)	84.5 (83.3–85.7)	80.3 (77.5–82.9)	82.8 (79.8–85.4)	85.6 (83.6–87.5)	75.3 (72.3–78.1)
Sensitivity (%)	81.4 (79.8–82.9)	76.1 (72.4–79.6)	78.8 (74.8–82.4)	81.6 (78.7–84.3)	72.4 (68.5–76.0)
Specificity (%)	90.1 (88.4–91.7)	88.5 (84.3–91.9)	90.5 (86.1–93.9)	92.1 (89.3–94.3)	80.5 (75.7–84.7)
AUC	0.93 (0.92–0.94)	0.90 (0.88–0.92)	0.91 (0.89–0.93)	0.94 (0.92–0.95)	0.82 (0.79–0.84)

AUC, area under the curve; CI, confidence interval.

<sup>a</sup>Optimal accuracy, sensitivity, and specificity cutoff values of 0.28 for all single gastric images, 0.66 for antrum, 0.38 for angularis, 0.27 for body, and 0.36 for fundus were used.

**Table 4.** Diagnostic characteristics for multiple images at different cutoff values

Cutoff value (probability score)	Accuracy (%)	Sensitivity (%)	Specificity (%)
0.1	86.1	97.4	61.3
0.2	89.6	94.8	78.2
0.3	92.5	92.9	91.5
0.4 <sup>a</sup>	93.8	91.6	98.6
0.5	91.8	88.1	100
0.6	87.2	81.3	100
0.7	82.3	74.2	100
0.8	78.5	68.7	100
0.9	67.3	52.3	100

<sup>a</sup>Optimal cutoff value was calculated using the Youden index.

(14). However, both studies used *H. pylori* serology as a reference standard which is not recommended in clinical practice. In our study, CNN using multiple gastric images (mean of  $8.3 \pm 3.3$  images per patient) demonstrated high sensitivity of 91.6% (95% CI 88.0%–94.4%), specificity of 98.6% (95% CI 95.0%–99.8%), and accuracy of 93.8% (95% CI 91.2%–95.8%). Furthermore, CNN using multiple gastric images demonstrated higher accuracy compared with any single gastric image or a single gastric image of the body. More importantly, CNN using multiple gastric images showed high sensitivity and specificity comparable with other direct *H. pylori* testing methods (histology with sensitivity of 88%–92% and specificity of 89%–98%, breath test with sensitivity of 96% and specificity of 93%, and stool antigen with sensitivity of 94% and specificity of 97%) recommended in clinical practice (15–17).

Our findings have clinical implications. For patients being evaluated for *H. pylori*-associated disease undergoing endoscopy, gastric biopsies are routinely obtained to perform a rapid urease test or histologic evaluation. Previous studies demonstrated that the evaluation of *H. pylori* according to endoscopic features is variable and inaccurate (2). Although magnification endoscopy or narrow band imaging can improve the diagnostic accuracy of *H. pylori* evaluation, the need for additional equipment and learning curve limit the generalizability in clinical practice (18,19). Therefore, evaluation of *H. pylori* using convolutional neural network that obviates the need for gastric biopsies while providing rapid diagnosis is appealing. Our results demonstrating high diagnostic accuracy suggest the possibility that CNN can replace gastric biopsies for evaluation of *H. pylori* infection. CNN as a disruptive technology can potentially reduce procedure time and utilization of resources. Although a higher diagnostic accuracy of *H. pylori* is needed for certain conditions (e.g., mucosa-associated lymphoid tissue lymphoma or peptic ulcer disease) associated with high disease morbidity, >90% diagnostic accuracy is likely acceptable for patients being evaluated for functional dyspepsia. Given the high prevalence of functional dyspepsia in the general population as the leading indication for *H. pylori* testing, our findings may be generalized to a large population undergoing endoscopy (20).

The strength of our study includes evaluation of *H. pylori* using CNN in a large patient population in a high prevalence area of *H. pylori* infection. Furthermore, all the patients in the study received immunohistochemistry examination of gastric biopsies, and a substantial portion (43%) received additional direct *H. pylori* testing to corroborate our reference standard.

Our study has limitations. Patients with peptic ulcers, an important indication for *H. pylori* evaluation, found during endoscopy were excluded which may limit the generalizability. High diagnostic accuracy of CNN demonstrated in our study will also require validation in other populations such as areas of low *H. pylori* prevalence. Furthermore, implementation of CNN in clinical practice will initially require large data for deep machine learning. In our study, more than 1,500 patients with nearly 11,000 gastric images were used to derive an accurate deep learning model. In addition, approximately 8 photographic documentations of gastric images were required to attain acceptable diagnostic property. Although CNN can effectively extract useful information and accurately classify images, the current technology for interpreting the categorized images is rudimentary (21–23). Further technical refinements including real-time assessment of *H. pylori* during live endoscopy rather than static images and compatibility with existing endoscopy platform will be vital for implementation in clinical practice. Improving the current speed of image classification (0.026 s/frame) will be important for real-time assessment of *H. pylori* infection during live endoscopy. Furthermore, future prospective studies comparing the diagnostic accuracy of convolutional neural network based on video recording vs gastric biopsies for evaluation of *H. pylori* infection using protocol gastric biopsies (antrum and body) and same-day C13 breath test as gold standard to eliminate the variability of gastric sampling in the current study will be invaluable. Finally, in addition to evaluation of *H. pylori* infection, we anticipate expanding applications of CNN during upper endoscopy to include the evaluation of gastric intestinal metaplasia and dysplasia (24).

In conclusion, CNN using multiple gastric images demonstrated high sensitivity, specificity, and accuracy for evaluating *H. pylori* infection, comparable with another recommended direct testing. Although our results are preliminary, CNN may potentially replace gastric biopsies in patients undergoing endoscopy for evaluation of *H. pylori*-associated diseases.

#### CONFLICTS OF INTEREST

**Guarantor of the article:** Weiling Hu, MD, PhD.

**Specific author contributions:** Wenfang Zheng and Xu Zhang jointly acted as first authors of this work. W.H., J.L., and J.S. made the conception and design; W.H., X.Z., G.Y., B.Y., J.W., S.L. and J.L. provided the endoscopic images; W.Z. and X.Z. independently screened and excluded unqualified images; J.L., X.Z. and T.Y. trained the CNN models; W.Z., J.L., J.J.K. and X.Z. did analysis and interpretation of the data; W.Z. and J.J.K. drafted the article.

**Financial support:** This work was supported by Public Welfare Research Project of Zhejiang Province (LGF18H160012), Medical Health Project of Zhejiang Province (2020RC064) and National Key Research and Development Program of China (2017YFC0113505).

**Potential competing interests:** None to report.

## Study Highlights

### WHAT IS KNOWN

- ✓ Endoscopy is frequently performed for evaluation of *H. pylori*-associated diseases.
- ✓ Evaluation of *H. pylori* at the time of endoscopy requires gastric biopsies as endoscopic impression alone is inaccurate.

### WHAT IS NEW HERE

- ✓ The Computer-Aided Decision Support System was developed to evaluate for *H. pylori* infection based on endoscopic images.

### TRANSLATIONAL IMPACT

- ✓ CNN may potentially replace gastric biopsies in patients undergoing evaluation for *H. pylori*-associated diseases.

## REFERENCES

1. Parsonnet J, Friedman GD, Vandersteen DP, et al. Helicobacter pylori infection and the risk of gastric carcinoma. *N Engl J Med* 1991;325:1127–31.
2. Laine L, Cohen H, Sloane R, et al. Interobserver agreement and predictive value of endoscopic findings for *H. pylori* and gastritis in normal volunteers. *Gastrointest Endosc* 1995;42:420–3.
3. Redeen S, Petersson F, Jönsson KA, et al. Relationship of gastroscopic features to histological findings in gastritis and Helicobacter pylori infection in a general population sample. *Endoscopy* 2003;35:946–50.
4. Togashi K. Applications of artificial intelligence to endoscopy practice: The view from JDDW 2018. *Dig Endosc* 2019;31:270–72.
5. Urban G, Tripathi P, Alkayali T, et al. Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy. *Gastroenterology* 2018;155:1069–78.e8.
6. Xiao J, Meng MQ. A deep convolutional neural network for bleeding detection in wireless capsule endoscopy images. *Conf Proc IEEE Eng Med Biol Soc* 2016;2016:639–42.
7. Zhu Y, Wang QC, Xu MD, et al. Application of convolutional neural network in the diagnosis of the invasion depth of gastric cancer based on conventional endoscopy. *Gastrointest Endosc* 2019;89:806–15.e1.
8. Cireşan D, Meier U, Schmidhuber J. Multi-column Deep Neural Networks for Image Classification. *Eprint Arxiv* 2012;157:3642–9.
9. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. *Communications of the ACM* 2017;60:84–90.
10. Malfertheiner P, Megraud F, O’Morain CA, et al. Management of Helicobacter pylori infection—the Maastricht V/Florence Consensus Report. *Gut* 2017;66:6–30.
11. Parkhi OM, Vedaldi A, Zisserman A. Deep face recognition. *BMVA Press* 2015;41:1–12.
12. Suzuki K. Overview of deep learning in medical imaging. *Radiol Phys Technol* 2017;10:257–73.
13. 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.
14. Shichijo S, Nomura S, Aoyama K, et al. Application of convolutional neural networks in the diagnosis of Helicobacter pylori infection based on endoscopic images. *EBioMedicine* 2017;25:106–11.
15. Laine L, Lewin DN, Naritoku W, et al. Prospective comparison of H&E, Giemsa, and Genta stains for the diagnosis of Helicobacter pylori. *Gastrointest Endosc* 1997;45:463–7.
16. Ferwana M, Abdulmajeed I, Alhajahmed A, et al. Accuracy of urea breath test in Helicobacter pylori infection: meta-analysis. *World J Gastroenterol* 2015;21:1305–14.
17. Gisbert JP, de la Morena F, Abaira V. Accuracy of monoclonal stool antigen test for the diagnosis of *H. pylori* infection: A systematic review and meta-analysis. *Am J Gastroenterol* 2006;101:1921–30.
18. Gonen C, Simssek I, Sarioglu S, et al. Comparison of high resolution magnifying endoscopy and standard videoendoscopy for the diagnosis of Helicobacter pylori gastritis in routine clinical practice: A prospective study. *Helicobacter* 2009;14:12–21.
19. Tahara T, Shibata T, Nakamura M, et al. Gastric mucosal pattern by using magnifying narrow-band imaging endoscopy clearly distinguishes histological and serological severity of chronic gastritis. *Gastrointest Endosc* 2009;70:246–53.
20. Miwa H, Kusano M, Arisawa T, et al. Evidence-based clinical practice guidelines for functional dyspepsia. *J Gastroenterol* 2015;50:125–39.
21. Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. *ECCV* 2014;8689:818–33.
22. Mahendran A, Vedaldi A. Understanding deep image representations by inverting them. *IEEE Computer Society* 2015; 5188–5196.
23. Kim I, Rajaraman S, Antani S. Visual interpretation of convolutional neural network predictions in classifying medical image modalities. *Diagnostics* 2019;9:38.
24. Cho BJ, Bang CS, Park SW, et al. Automated classification of gastric neoplasms in endoscopic images using a convolutional neural network. *Endoscopy* 2019.

**Open Access** This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.