

Deep learning analyzes *Helicobacter pylori* infection by upper gastrointestinal endoscopy images



Authors

Takumi Itoh¹, Hiroshi Kawahira^{2,3}, Hirotaka Nakashima⁴, Noriko Yata⁵

Institutions

- 1 Department of Medical System Engineering, Graduate School of Engineering, Chiba University
- 2 Center for Frontier Medical Engineering, Chiba University, Chiba, Japan
- 3 Department of Frontier Surgery, Graduate School of Medicine, Chiba University, Chiba, Japan
- 4 Department of Gastroenterology, Foundation for Detection of Early Gastric Carcinoma, Tokyo, Japan
- 5 Department of Information Processing and Computer Science, Graduate School of Advanced Integration Science, Chiba University, Chiba, Japan

submitted 12.9.2017

accepted after revision 22.9.2017

Bibliography

DOI <https://doi.org/10.1055/s-0043-120830> |
Endoscopy International Open 2018; 06: E139–E144
© Georg Thieme Verlag KG Stuttgart · New York
ISSN 2364-3722

Corresponding author

Hiroshi Kawahira, MD, PhD, Chiba University, Center for Frontier Medical Engineering, 1-33, Yayoi-cho, Inage-ku, Chiba 263-8522, Japan
Fax: +81432903124
hk@faculty.chiba-u.jp

ABSTRACT

Background and study aims *Helicobacter pylori* (HP)-associated chronic gastritis can cause mucosal atrophy and intestinal metaplasia, both of which increase the risk of gastric cancer. The accurate diagnosis of HP infection during routine medical checks is important. We aimed to develop a convolutional neural network (CNN), which is a machine-learning algorithm similar to deep learning, capable of recognizing specific features of gastric endoscopy images. The goal behind developing such a system was to detect HP infection early, thus preventing gastric cancer.

Patients and methods For the development of the CNN, we used 179 upper gastrointestinal endoscopy images obtained from 139 patients (65 were HP-positive: ≥ 10 U/mL and 74 were HP-negative: < 3 U/mL on HP IgG antibody assessment). Of the 179 images, 149 were used as training images, and the remaining 30 (15 from HP-negative patients and 15 from HP-positive patients) were set aside to be used as test images. The 149 training images were subjected to data augmentation, which yielded 596 images. We used the CNN to create a learning tool that would recognize HP infection and assessed the decision accuracy of the CNN with the 30 test images by calculating the sensitivity, specificity, and area under the receiver operating characteristic (ROC) curve (AUC).

Results The sensitivity and specificity of the CNN for the detection of HP infection were 86.7% and 86.7%, respectively, and the AUC was 0.956.

Conclusions CNN-aided diagnosis of HP infection seems feasible and is expected to facilitate and improve diagnosis during health check-ups.

Introduction

A strong link between *Helicobacter pylori* (HP) infection and gastric cancer has been reported [1, 2]. HP is the leading cause of HP infection-associated gastritis and can cause chronic gastritis, gastroduodenal ulceration, mucosal atrophy, and intestinal metaplasia [3]. The latter 2 conditions are known risk factors for the development of gastric cancer [1, 4]. Eradication of HP is known to improve gastric mucosal atrophy and inhibit the development of intestinal metaplasia [5]. Thus, it is important

to diagnose HP infection to avoid the potential development of gastric cancer. We are concerned with the accurate diagnosis of HP infection during routine medical check-ups.

Using standard endoscopy, HP infection is diagnosed on the basis of gastric mucosal redness and swelling [6]; however, even this approach requires advanced skills and knowledge [4], and several years of training are necessary for endoscopists to attain the necessary diagnostic expertise [7]. Machine learning can be applied to overcome the problems of diagnosis, and a convolutional neural network (CNN) optimized for the diag-

nosis of HP infection may be clinically beneficial in preventing the development of gastric cancer. Machine learning is a method of data analysis that allows the discovery of specific patterns in large datasets. Deep learning is a type of machine learning that is based on a set of algorithms that attempt to model high-level abstractions in data. It is a multilayered approach that imitates cerebral neural networks and uses various layers to automatically extract features from images or voices. A CNN can be trained to automatically extract image features and then recognize patterns after multilayered learning of image data achieved through deep learning [8]. A CNN is similar in structure to a neocognitron, which is an image recognition system derived from computational neuroscience [8]. One important characteristic of a CNN is that interhierarchy operations can be stated as convolution operations. Thus, a CNN exhibits high accuracy when used for recognition of images and voice.

Aiming to simplify endoscopic screening for HP infection, we constructed a CNN that was optimized to diagnose HP infection by learning endoscopic images. Caffe was used as the framework for the CNN [9]. In the present study, we used a CNN designed for generic object recognition and then used a fine-tuning strategy to transfer the recognition capabilities of the CNN to endoscopic images, to further aid in the diagnosis of HP infection. The ultimate goal of the development of this system was the early detection of HP infection, thus, preventing gastric cancer.

Patients and methods

Preparation and experimental data

This prospective, cohort study was approved by the ethics committee of the Foundation for the Detection of Early Gastric Carcinoma (approval No.15-02). The study included white-light endoscopic images that had been obtained from 139 individuals during annual company-sponsored health check-ups. As this study was exploratory study, sample size was determined according to practicability for sample collection and analysis. We referred to the papers related to previously reported machine learning[10]. All endoscopic examinations were performed with an EG-L580NW endoscope (Fujifilm, Tokyo, Japan) by the same doctor (H.N.), certified by the board of the Japan Gastroenterological Endoscopy Society.

All 139 individuals provided their written consent for an HP blood test. The distributions of clinical diagnoses are indicated based on the degree of mucosal atrophy according to the Kimura and Takemoto classification in ► **Table 1** [11]. Blood was drawn from each individual, and the serum was tested for HP IgG antibodies. An antibody titer of ≥ 10 U/mL was considered positive for HP infection and a titer of < 3 U/mL was considered negative. To avoid the inclusion of false-negative test results, individuals with a serum antibody titer of ≥ 3 U/mL and ≤ 9 U/mL and individuals who underwent HP eradication were excluded from the study to improve the diagnosis of the CNN. Of the 139 individuals tested, 65 were positive for HP infection, whereas 74 showed negative results.

From the 139 individuals, we obtained 179 endoscopic images of the lesser curvature of the stomach. For machine learn-

► **Table 1** Patient numbers for clinical diagnosis of gastritis.

	Training	Test	Pt. number
Non	45	14	59
C-1	9	1	10
C-2	9	3	12
C-3	6	2	8
O-1	24	5	29
O-2	14	4	18
O-3	2	1	3
Total	109	30	139

ing, 149 of the 179 images were used, and the remaining 30 were set aside to be used as test images. These 30 images were obtained from 15 patients who tested positive for HP infection and 15 patients who tested negative. Representative HP-positive and HP-negative images are shown in ► **Fig. 1**. The 149 learning images were large batch images of 800×800 pixels each (► **Fig. 2**). Processing of these images was performed at angles of 45° , 90° , and 180° for data augmentation, yielding a total of 596 images for learning (► **Table 2**).

Machine learning

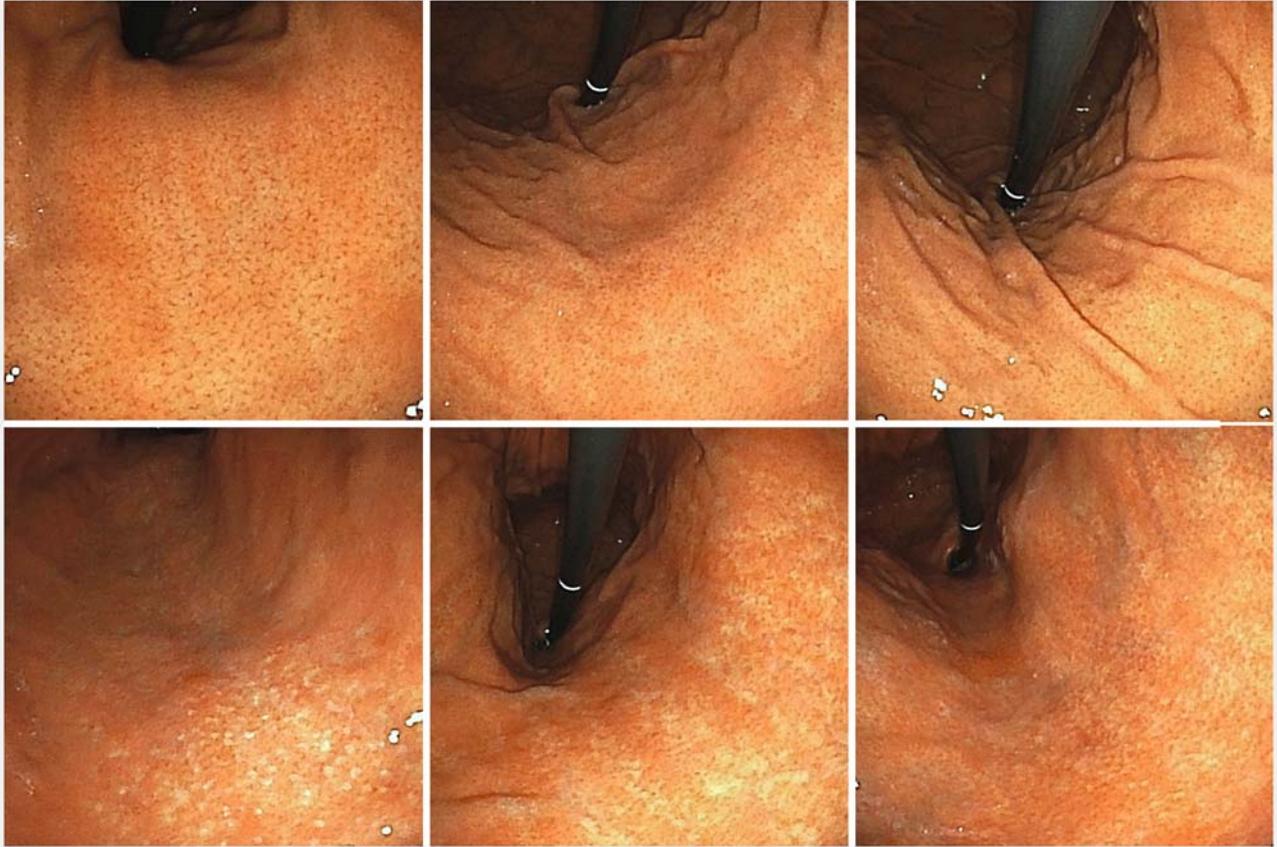
We used GoogLeNet DCNN pretuned for generic object recognition[12]. GoogLeNet, developed by Google, is a 22-layer network that won acclaim during the ImageNet Large Scale Visual Recognition Competition (ILSVRC) [13, 14]. A diagram of CNN learning flow is presented in ► **Fig. 3**.

In general, CNN learning has strong early-stage dependencies. There were limited images available for the learning process; however, transfer learning effectively leveraged the available dataset. The transfer learning process was based on fine-tuning [15], which was based on a pretrained network, and the parameters used were early values. Furthermore, transfer learning was the technique used for learning the task data set [16]. Flow images, like the identification layer, were the only factors added to the task data set, and we trained the network's weights and biases so that the network's output would correctly identify the input image.

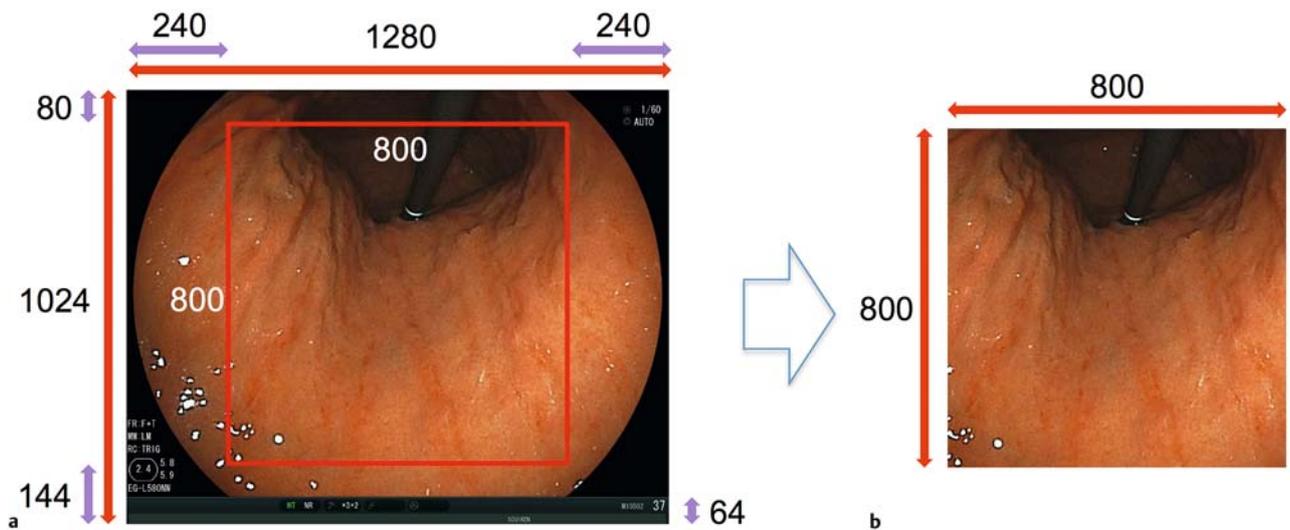
Stochastic gradient descent was used to optimize the network [17]. The momentum coefficient for the rate of learning, batch size, number of iterations, and attenuation were 0.0005, 20, 20,000, and 0.0005, respectively.

Creating a learning device by means of GoogLeNet

As discussed above, presence or absence of HP infection was confirmed according to the serum HP IgG antibody levels in the patients from whom the 149 endoscopic images were obtained. We compared the results of the serum HP IgG antibody tests with the results obtained by applying our new machine learning algorithm and investigated the sensitivity, specificity, and area under the receiver operating characteristic (ROC) curve (AUC) for our system.



► **Fig. 1** Examples of endoscopic images obtained from individuals who, upon laboratory tests, were shown to be negative (upper row) or positive (lower row) for HP infection.



► **Fig. 2** Image processing by means of deep learning focused on the center of the image. Image resolution was 800 x 800 pixels. **a** Before processing. **b** After processing.

► **Table 2** Breakdown of training images and test images.

	HP infection status	No. of endoscopic images	No. of images after data augmentation
Training images	Positive	70	280
	Negative	79	316
Test images	Positive	15	–
	Negative	15	–

Evaluation

As aforementioned, 30 images (15 HP-positive and 15 HP-negative images) were used as test images. CNN decision accuracy, that is, screening accuracy, was determined on the basis of the ROC curve, derived by plotting sensitivity against specificity, for which we calculated the AUC. No cut-off values (values differentiating between positive and negative results) were used, and the ROC curve was obtained when we plotted the values. Values ranged from 0 to 1, and values closer to 1 indicated high classification accuracy.

Results

When we considered CNN output values of 0.5–1.0 as positive and values of 0–0.49 as negative, sensitivity was 86.7% and specificity was 86.7%. The AUC was 0.956 (► **Fig. 4**).

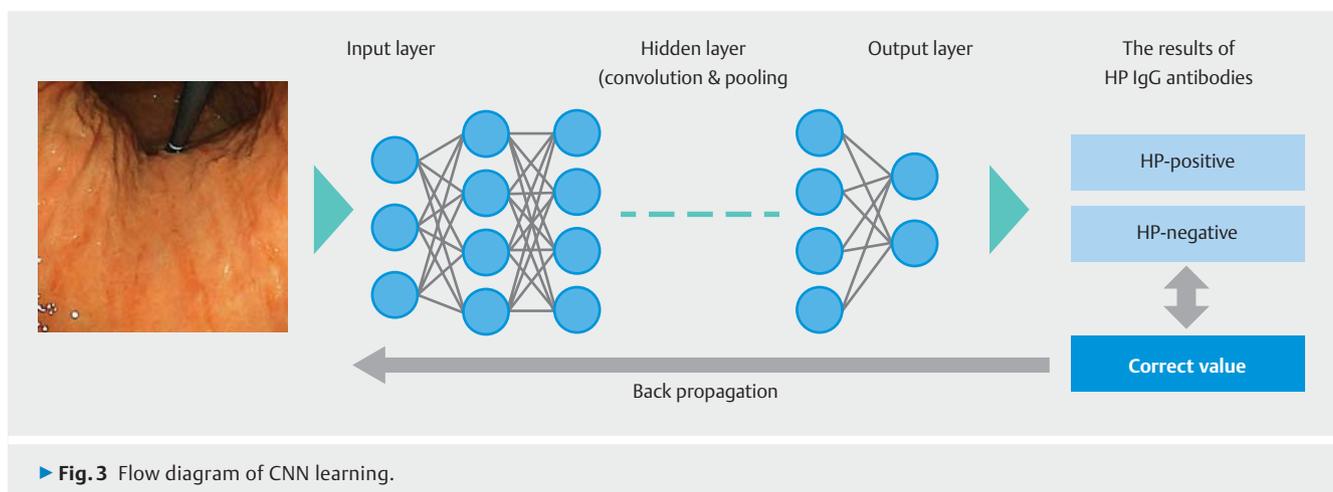
Discussion

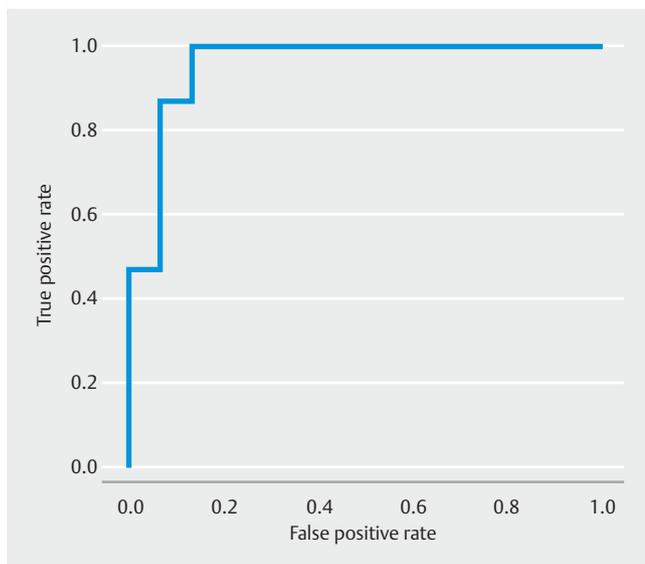
Diagnosing cancer early and providing appropriate treatment are vital. Mucosal atrophy and intestinal metaplasia that can occur with HP infection are important factors in the development of gastric cancer [2–6, 18, 19]. If endoscopy is performed at the time of an individual's annual health check-up, the five key Kyoto classification features that indicate a person's risk of gastric cancer (atrophy, intestinal metaplasia, enlargement and tortuosity of the gastric folds, nodularity, and diffuse reddening) can be examined; however, these features are not easily identified,

and all endoscopists do not possess the same diagnostic ability [20].

Endoscopic evidence of HP infection is, however, rapidly obtained, and it results in an accurate diagnosis. Ji et al. noted the relative difficulty of real-time diagnosis under white light and reported the effectiveness of magnifying endoscopy with narrow band imaging (NBI) [21]. Nodularity recognized under white-light endoscopic examination is highly specific for HP infection (96% specificity), but discovery of such nodularity is not sufficiently sensitive for correctly identifying patients without HP infection (32% sensitivity) [21]. In general, the diagnostic ability increases under white light; therefore, chromoendoscopy, that is, dyeing tissues and viewing them endoscopically, is useful. Magnifying endoscopy allows for detailed observation of the mucosal structure, leading to the diagnosis of HP infection. Diagnosis is facilitated by observing and classifying features such as hypertrophic mucosa and absence of round pits with regularly arranged collecting venules; however, achieving an objective diagnosis is difficult [22, 23], making NBI necessary. HP infection diagnosed using NBI is classified on the basis of the following three abnormal pit patterns: type 1 (slightly enlarged, round pits with unclear or irregular subepithelial capillary networks [SECNs]); type 2 (obviously enlarged oval or elongated pits with an increased density of irregular vessels); and type 3 (well-demarcated oval or tubulovillous pits with clearly visible coiled or wavy vessels). The reported sensitivity and specificity of these pit patterns for the detection of HP infection are 95.2% and 82.2% [24]. Both magnifying endoscopy and NBI are required to accurately identify the absence of mucosal pits and collecting venules as well as the presence of irregular SECNs; thus, diagnostic skill is necessary.

Diagnosis of HP infection relies on the results of endoscopic rapid urease testing (RUT), endoscopic biopsy and culture, the measurement of serum antibodies, the results of the urea breath test (UBT), and the results of stool antibody testing. Endoscopy-based tests, such as RUT or pathological diagnosis from biopsy samples, can yield false-negative results if mucosa without HP is collected. RUT is effective; however, but we did not perform a biopsy in all cases during endoscopic examination. In addition, RUT was considered expensive and hence not





► Fig. 4 Receiver-operating characteristic curve.

performed. Serum antibody tests always yield false-negative results in small children immediately after infection, and false-negative results are possible within a few months after eradication therapy. The test device used for the UBT is not commonly available. Stool antigen testing is highly sensitive and specific and can be used for diagnosis even in children. Nevertheless, collection and handling of the specimens are difficult. Serum antibody testing is recommended in Japan to screen for HP infection if endoscopy is not performed [25].

The UBT, which does not require endoscopy, has a sensitivity and specificity of 97% and 97%, respectively, and the serum antibody assay has a sensitivity and specificity of 97.7% and 95.6%, respectively [26,27]. UBT is a very useful and sensitive method, the serological *H. pylori* antibody test is the gold standard test to detect *H. pylori* infection [25].

We developed and tested a CNN system to be used as a diagnostic aid. The sensitivity and specificity of the CNN system were 86.7% and 86.7%, respectively, and the AUC was 0.956. The AUC was large, and we believe that our system was capable of consistent recognition. The sensitivity and specificity of the UBT and serum antibody testing in this study differed by 10% in terms of accuracy, so increasing precision will be the next challenge. The CNN system developed to aid in the diagnosis of HP gastritis is not influenced by the site at which the endoscopic biopsy specimen is obtained. In fact, no biopsy specimens are needed. The system is advantageous in that the diagnostic capabilities of endoscopy specific to HP can be automated.

Watanabe et al. reported that it was difficult to distinguish between *H. pylori*-infected and *H. pylori*-eradicated patients [7]. In this study, we excluded patients with a history of *H. pylori* eradication or with serum antibody titers of ≥ 3 U/ml and ≤ 9 U/ml. It is important to distinguish *H. pylori*-eradicated, -infected and -uninfected patients, and accurate training data are necessary to create a CNN for the diagnosis of *H. pylori* infection. Patients with *H. pylori* IgG antibody titers of ≥ 3 U/mL and ≤ 9 U/mL were considered to be included within a serological boundary region

of *H. pylori* infection, which the training data that we used for the CNN did not contain. Including this in the training data set would improve the diagnostic accuracy of the CNN. We understand that the analysis in this study was quite different from clinical practice. To improve the CNN, we plan to prospectively evaluate its diagnostic accuracy in patients after eradication of *H. pylori* so that *H. pylori*-positive, -negative, and -eradicated diagnoses are included in the CNN.

Accurately diagnosing HP infection using standard endoscopy is difficult. Bah et al. conducted a prospective investigation of the diagnostic accuracy based on gastric endoscopy images. The resulting sensitivity and specificity were 75% and 63%, respectively, and the authors concluded that it is difficult to diagnose HP gastritis from endoscopic images alone [28]. Their findings reflect the fact that diagnostic skill, and thus accuracy, varies from one endoscopist to another. We have two reasons for selecting images of the lesser curvature of the stomach. Firstly, the lesser curvature has a higher diagnostic sensitivity and specificity than the greater curvature (data not shown). Secondly, we wanted to make the diagnosis of *H. pylori* infection as simple as possible. A system to aid and support diagnosis, whether performed by residents or experts, is desirable and will be very useful. In this study, we hypothesized atrophy in the corpus lesser curvature is an endoscopic indicator for *H. pylori* infection. However, in the Western countries, HP gastritis confined to the antrum and does not progress to the corpus much [29]. Therefore, this algorithm may not be so useful for patients in non-East Asian countries. An optimized CNN can be used as an automated diagnostic aid to identify HP infection. The developed CNN system will allow the diagnosis of HP infection at an early stage and facilitate appropriate timing of eradication. Therefore, it may be possible to reduce the risk of gastric cancer with this system.

Conclusions

We developed a CNN system to aid in the endoscopic diagnosis of HP infection. Good results were obtained, suggesting that the system is useful for identification of HP infection. To the best of our knowledge, this is the first report of such a CNN. We believe that this machine learning system will effectively support clinical diagnosis irrespective of whether the system is used by residents or experts. Its use may help ease the diagnostic burden of physicians. By clarifying patient risks, it may also help in the identification of appropriate treatment strategies. In the future, we plan to aim for identification accuracy similar to that in the present study and to advance the parameter adjustment and data augmentation.

Competing interests

None

References

- [1] Tahara T, Arisawa T, Shibata T et al. Risk prediction of gastric cancer by analysis of aberrant DNA methylation in non-neoplastic gastric epithelium. *Digestion* 2007; 75: 54–61
- [2] Uemura N, Okamoto S, Yamamoto S et al. Helicobacter pylori infection and the development of gastric cancer. *N Engl J Med* 2001; 345: 784–789
- [3] Goodwin CS. Helicobacter pylori gastritis, peptic ulcer, and gastric cancer: clinical and molecular aspects. *Clin Infect Dis* 1997; 25: 1017–1019
- [4] Nomura S, Terao S, Adachi K et al. Endoscopic diagnosis of gastric mucosal activity and inflammation. *Dig Endosc* 2013; 25: 136–146
- [5] Tahara T, Shibata T, Horiguchi N et al. A possible link between gastric mucosal atrophy and gastric cancer after Helicobacter pylori eradication. *PLoS One* 2016; 11: e0163700
- [6] 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: 420–423
- [7] Watanabe K, Nagata N, Shimbo T et al. Accuracy of endoscopic diagnosis of Helicobacter pylori infection according to level of endoscopic experience and the effect of training. *BMC Gastroenterol* 2013; 13: 128–134
- [8] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst (NIPS Proceedings)* 2012: 1106–1114
- [9] Jia Y, Shelhamer E, Donahue J et al. Caffe: convolutional architecture for fast feature embedding. *Proc ACM Int Conf Multimed* 2014: 675–678
- [10] Mori Y, Kudo S, Wakamura K et al. Novel computer-aided diagnostic system for colorectal lesions by using endocytoscopy (with videos). *Gastrointest Endosc* 2015; 81: 621–629
- [11] Kimura K, Takemoto T. An endoscopic recognition of the atrophic border and its significance in chronic gastritis. *Endoscopy* 1969; 1: 87–97
- [12] Szegedy C, Liu W, Jia Y et al. Going deeper with convolutions. 2015 IEEE Conf Comput Vis Pattern Recognit 2015: 1–9
- [13] Deng J, Dong W, Socher R et al. ImageNet: a large-scale hierarchical image database. 2009 IEEE Conf Comput Vis Pattern Recognit 2009: 2–9
- [14] Russakovsky O, Deng J, Su H et al. ImageNet Large Scale Visual Recognition Challenge. *Int J Comput Vis* 2015; 115: 211–252
- [15] Girshick R, Donahue J, Darrell T et al. Rich feature hierarchies for accurate object detection and semantic segmentation. *Proc IEEE Comput Conf Comput Vis Pattern Recognit* 2014: 580–587
- [16] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. *Science* 2006; 313: 504–507
- [17] Bottou L. Large-scale machine learning with stochastic gradient descent. *Proc COMPSTAT'2010* 2010: 177–186
- [18] Bornschein J, Malfertheiner P. Gastric carcinogenesis. *Langenbecks Arch Surg* 2011; 396: 729–742
- [19] Asaka M, Sugiyama T, Nobuta A et al. Atrophic gastritis and intestinal metaplasia in Japan: results of a large multicenter study. *Helicobacter* 2001; 6: 294–299
- [20] Sugano K, Tack J, Kuipers EJ et al. faculty members of Kyoto Global Consensus Conference. Kyoto global consensus report on Helicobacter pylori gastritis. *Gut* 2015; 64: 1353–1367
- [21] Ji R, Li YQ. Diagnosing Helicobacter pylori infection in vivo by novel endoscopic techniques. *World J Gastroenterol* 2014; 20: 9314–9320
- [22] Okubo M, Tahara T, Shibata T et al. Usefulness of magnifying narrow-band imaging endoscopy in the Helicobacter pylori-related chronic gastritis. *Digestion* 2011; 83: 161–166
- [23] Nakagawa S, Kato M, Shimizu Y et al. Relationship between histopathologic gastritis and mucosal microvasculature: Observations with magnifying endoscopy. *Gastrointest Endosc* 2003; 58: 71–75
- [24] 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: 246–253
- [25] Lee S-Y. Current progress toward eradicating Helicobacter pylori in East Asian countries: differences in the 2013 revised guidelines between China, Japan, and South Korea. *World J Gastroenterol* 2014; 20: 1493–1502
- [26] Di Rienzo TA, D'Angelo G, Ojetti V et al. 13C-Urea breath test for the diagnosis of Helicobacter pylori infection. *Eur Rev Med Pharmacol Sci* 2013; 17: 51–58
- [27] Kato M, Asaka M, Saito M et al. Clinical usefulness of urine-based enzyme-linked immunosorbent assay for detection of antibody to Helicobacter pylori: a collaborative study in nine medical institutions in Japan. *Helicobacter* 2000; 5: 109–119
- [28] Bah A, Saraga E, Armstrong D et al. Endoscopic Features of Helicobacter pylori-related gastritis. *Endoscopy* 1995; 27: 59–56
- [29] Naylor GM, Gotoda T, Dixon M et al. Why does Japan have a high incidence of gastric cancer? Comparison of gastritis between UK and Japanese patients *Gut* 2006; 55: 1545–1552