JGH Open An open access journal of gastroenterology and hepatology Open Access
JGHF Journal of Gastroenterology and Hepatology Foundation

doi:10.1002/jgh3.12995

ORIGINAL ARTICLE

Diagnostic usefulness of deep learning methods for *Helicobacter pylori* infection using esophagogastroduodenoscopy images

Daesung Kang,* Kayoung Lee[†] and Jinseung Kim[†]

*Department of Healthcare Information Technology, Inje University, Gimhae and [†]Department of Family medicine, Busan Paik Hospital, Inje University College of Medicine, Busan, Republic of Korea

Key words
artificial intelligence, deep learning, *Helicobacter pylori*.

Accepted for publication 23 October 2023.

Correspondence
Jinseung Kim, Department of Family medicine, Busan Paik Hospital, Inje University College of Medicine, 75, Bokji-ro, Busanjin-gu, Busan 47392, Republic of Korea.
Email: jinseungkim@inje.ac.kr

Declaration of conflict of interest: The authors declare no conflict of interest.
Author contribution: All authors contributed to the study conception and design. Jinseung Kim, Daesung Kang, and Kayoung Lee contributed to material preparation, data collection, and analysis. Jinseung Kim wrote the first draft of the manuscript, and all authors commented on its previous versions. All authors have read and approved the final manuscript.
Financial support: This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HR21C1003)

Abstract

Background and Aims: We aimed to assess the diagnostic potential of deep convolutional neural networks (DCNNs) for detecting *Helicobacter pylori* infection in patients who underwent esophagogastroduodenoscopy and *Campylobacter*-like organism tests.
Methods: We categorized a total of 13,071 images of various gastric sub-areas and employed five pretrained DCNN architectures: ResNet-101, Xception, Inception-v3, InceptionResnet-v2, and DenseNet-201. Additionally, we created an ensemble model by combining the output probabilities of the best models. We used images of different sub-areas of the stomach for training and evaluated the performance of our models. The diagnostic metrics assessed included area under the curve (AUC), specificity, accuracy, positive predictive value, and negative predictive value.
Results: When training included images from all sub-areas of the stomach, our ensemble model demonstrated the highest AUC (0.867), with specificity at 78.44%, accuracy at 80.28%, positive predictive value at 82.66%, and negative predictive value at 77.37%. Significant differences were observed in AUC between the ensemble model and the individual DCNN models. When training utilized images from each sub-area separately, the AUC values for the antrum, cardia and fundus, lower body greater curvature and lesser curvature, and upper body greater curvature and lesser curvature regions were 0.842, 0.826, 0.718, and 0.858, respectively, when the ensemble model was used.
Conclusions: Our study demonstrates that the DCNN model, designed for automated image analysis, holds promise for the evaluation and diagnosis of *Helicobacter pylori* infection.

Introduction

Helicobacter pylori (*H. pylori*) is one of the most common infectious diseases worldwide, with a prevalence of 11% in Northern Europe, 23.1% in Canada, and 30% in the United States, and up to 72–82% in South America and 91% in Nigeria[1] In South Korea, prevalence of up to 50% has been reported.[2] *H. pylori* infection is associated with gastric adenocarcinoma, peptic ulcer, gastritis, mucosal-associated lymphoid tissue lymphoma, and immune thrombocytopenic purpura.[3,4] Furthermore, chronic atrophic gastritis, intestinal metaplasia in superficial gastritis, dysplasia, and progression of gastric cancer have been found to be strongly associated with *H. pylori* infection.[5,6]

With over 1 million new cases and an estimated 769 000 deaths (equivalent to 1 in every 13 deaths) occurring in 2020, gastric cancer has the fifth highest incidence among cancers and is the fourth leading cause of cancer death. Incidence rates are twofold higher for men than women.[7] Chronic *H. pylori* infection is the most common cause of non-cardia stomach cancer, accounting for nearly all cases.[8,9] The International Agency for Research on Cancer has classified *H. pylori* as a group 1 carcinogen based on the results of previous studies. The development of gastric cancer following *H. pylori* infection is likely to occur via an indirect effect of *H. pylori* on gastric epithelial cells by generating inflammation and a direct effect on epithelial cells. *H. pylori* may also directly regulate epithelial cell activity via

© 2023 The Authors. *JGH Open published by Journal of Gastroenterology and Hepatology* Foundation and John Wiley & Sons Australia, Ltd.
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

bacterial compounds such as cytotoxin-associated gene A. The Asia–Pacific Gastric Cancer Consensus recommends screening and treatment of *H. pylori* infection in communities with a high incidence of gastric cancer as an effective strategy for prevention of this disease.[10]

Atrophy, mucosal edema, enlarged mucosal folds, diffuse erythema, and mucosal nodularity are the characteristic features of *H. pylori*-induced gastritis. Conversely, fundic gland polyps and regular arrangements of collecting venules are predictive indicators of *H. pylori*-naïve stomach.[11,12] These endoscopic changes are not objective indicators because of the possible intra- or inter-observer variability in the optical diagnosis of *H. pylori*-infected mucosa.[13] Recent developments in analytical methods, such as neocognitron, support vector machine, and deep learning, have resulted in advances in artificial intelligence (AI) technology.[14] As part of a broad family of machine-learning methods based on learning data representations, deep learning architectures are particularly suitable for qualifying images and are known for their high performance in detection, classification, and segmentation.[15] AI using deep learning algorithms has already surpassed human-level image recognition and is increasingly used in clinical practice for image recognition and classification.[16,17] AI is expected to show promising diagnostic performance in detecting cancer or neoplastic lesions using endoscopic imaging and in classifying neoplastic or non-neoplastic lesions in the gastrointestinal tract.[18]

Therefore, we aimed to evaluate the diagnostic usefulness of different deep learning approaches in detecting *H. pylori* infection, as well as the diagnostic usefulness of the deep learning methodology based on the anatomic region of the stomach.

Materials and methods

Study design. This study was performed in a single hospital. Image data from 1268 patients who underwent esophagogastroduodenoscopy (EGDS) and *Campylobacter*-like organism (CLO) tests to confirm *H. pylori* infection without prior gastric surgery, procedures, or anatomical abnormalities were obtained from 2590 (positive: 1261, negative: 1329) patients who underwent EGDS and CLO tests simultaneously at Busan Paik Hospital between January 2019 and April 2020. A commercially available CLO kit (Pylo Plus; Gulf Coast Scientific, Oldsmar, FL, USA) was used. EGDS was performed at Busan Paik Hospital, and images were captured using a standard endoscope (GIF-H260; Olympus Medical Systems, Co., Ltd., Tokyo, Japan) and standard endoscopic video system (EVIS LUCERA ELITE CV-290/CLV-290SL; Olympus Medical Systems). Of the participants, 809 had tested CLO-negative and 459 CLO-positive. This study included 13 071 images, of which 7279 were negative and 5792 were positive for the CLO test. The endoscopist visually classified the images, and the images of the gastric mucosa were grouped individually. The angle, antrum, cardia, fundus, lower body lesser curvature (LC), lower body greater curvature (GC), upper body LC, and upper body GC were also categorized.

Cases that displayed substandard video quality, characterized by either blurring or being out of focus, or those that presented suspected lesions specific to the mucosal area, such as polyps, hemorrhages, ulcers, or cancer, were eliminated from consideration. As *H. pylori* is not evenly distributed throughout the gastric mucosa, it is recommended that tissues from the antrum and body be collected and examined together to improve the diagnostic rate.[19] Tissue samples were obtained from the antrum and body mucosa and placed in the test kit. A positive CLO was determined by a change in color from yellow to red within 60 min of the sample being deposited in the kit at room temperature, as recommended by the manufacturer. This study was approved by the Inje University Busan Paik Hospital's Institutional Review Board (IRB) in Busan, Korea (IRB approval number: 2021-08-059), which waived the requirement for written informed consent.

Image preprocessing. The raw EGDS images include black regions that contain patient information. In Figure 1a, patient information is pictorially covered by a red rectangular box to prevent leakage of personal information. In addition to protecting patient privacy, black areas should be removed to prevent deep learning models from mistraining. Therefore, image processing techniques were applied to remove unnecessary black regions from each EGDS image in the dataset to minimize human intervention and to obtain consistent images.

To obtain a region of interest (ROI) from an image, we used only the red channel in the red, green, blue (RGB) image because the EGDS images are rich in red color. A Canny edge was applied to find the edge in this grayscale image, as shown in Figure 1b. After dilation of the morphological operation on the edge image, image filling was performed to obtain the results shown in Figure 1c. The image was complemented and dilated to remove personal information. Consequently, the image shown in Figure 1d was obtained. After complementing the resulting image again, the coordinate information of the four red dots at the corner of the white area, which is shown in Figure 1e, was extracted. An ROI image, as shown in Figure 1f, was obtained using the extracted coordinate information. Through these preprocessing steps, we generated an appropriate image for inputting into the convolutional neural network (CNN).

Deep convolutional neural network. We used five pretrained deep convolutional neural network (DCNN) architectures: ResNet-101,[20] Xception,[21] Inception-v3,[22] InceptionResnet-v2,[23] and DenseNet-201.[24] The networks had already learned to extract informative features from the ImageNet dataset. Five pretrained DCNN models were used to discriminate *H. pylori* infection by replacing the classifier layer with a new classifier. This was performed by initializing the weight of the convolution layers with the pretrained weights of the pretrained DCNN models while initializing the new classifier layers with random weights. The EGDS image was then used to fine-tune the parameters of the convolution and classifier layers. Fine-tuning of the five DCNN models was implemented in MATLAB 2020a using the deep-learning toolbox on a Windows 10 PC with an NVIDIA RTX 2080Ti GPU. The Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$) was used with initial learning rates of 5e-4, 1e-4, and 5e-5 to fine-tune the pretrained networks. We applied L2 regularization (weight decay) to penalize large weights to avoid overfitting and to generalize the results well. The regularization hyperparameter,

© 2023 The Authors. *JGH Open published by Journal of Gastroenterology and Hepatology* Foundation and John Wiley & Sons Australia, Ltd.

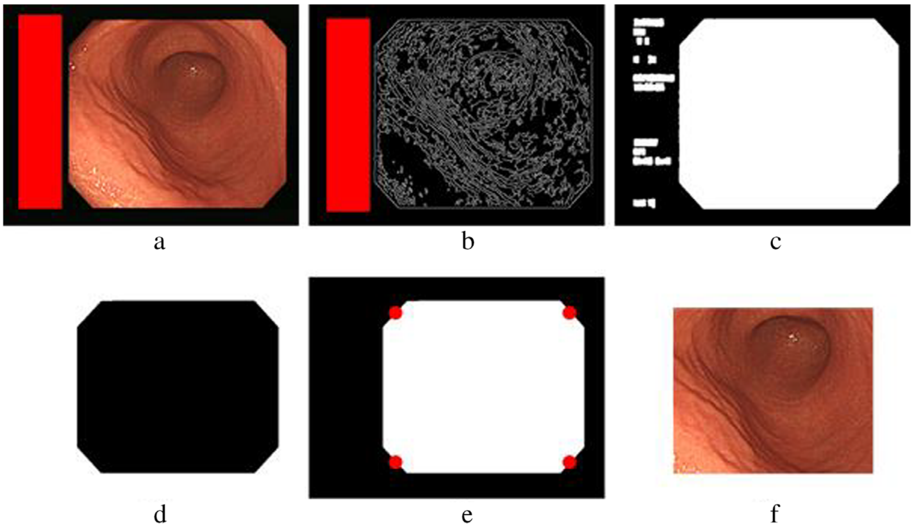

Figure 1 Procedure to remove the black region of an image using image processing techniques. (a) Image captured by the endoscopic system. The patient information in (a) is pictorially covered by a red rectangular box. (b) image in which the Canny edge algorithm is applied to the red channel of the color image (a). Morphological operations such as dilation, filling, and complement were used in (c) and (d) to identify corner coordinates in (e) marked with red dots. Using the corner coordinates in (e), the region of interest as shown in (f) was cropped.

which controls the amount of regularization, was set to 5e-4 and 1e-3 for all DCNN models. Furthermore, we adopted an early stopping strategy to monitor the validation loss with a patience set of 30 epochs to prevent overfitting. The number of epochs to wait for a lower loss before aborting if no progress was noted in the validation set. The sizes of the mini-batch were set to 32 for the Xception and Inception-v3 models, 16 for the ResNet-101 and InceptionResnet-v2 models, and 8 for the DenseNet-201 model. The mini-batch size was determined based on the maximum capacity of the GPU. The best DCNN model with the best AUC in the validation set was selected. To improve discrimination power, we adopted an ensemble model that constructs different DCNN models. The ensemble model was obtained by averaging the output probabilities of the best DCNN models.

Data configuration. Of the 13 071 EGDS images, 8894 images (positive 4953, negative 3941) were randomly chosen as training data and 1570 images (positive 874 and negative 696) as validation data. The remaining 2607 images (positive 1452 and negative 1155) were independently chosen as test data. After image preprocessing, as depicted in Figure 1, we applied the following data augmentation methods to the images to reduce overfitting and increase the number of training examples: resizing the input image using a scale factor selected randomly from the range (0.8, 1.2); flipping with 50% probability in horizontal and vertical axes; rotating by an angle selected randomly from the range (−20, 20) degrees; and shifting horizontally and vertically by a distance selected randomly from the range (−30, 30) pixels. Then, the input images were resized to 224 × 224 pixels for the ResNet-50 and DenseNet-201 models, and to 299 × 299 pixels for the Xception, Inception-v3, and Inceptionresnet-v2 models to be compatible with the proposed network. Before being fed to the training networks, an input image was normalized by channel-wise mean subtraction and divided by the standard deviation using the ImageNet mean and standard deviation values, respectively. The validation and test data were not augmented.

Visual explanation by Grad-CAM. We employed a gradient-weighted class activation mapping (Grad-CAM) technique on the classified images to understand why the DCNN model made its decisions.[25] The Grad-CAM method utilizes the gradient of the classification score with respect to the convolutional features determined by the network to provide a visual explanation of the areas of the image that influence the classification of the model. The areas where this gradient is large are the areas where the final score depends the most on the data. Grad-CAM visualizes the resulting image through a heatmap, where red typically represents high values and blue represents low values. Figures 2 and 3 show several EGDS images and their respective heatmaps generated by Grad-CAM using the InceptionResnet-v2 model. The Grad-CAM heatmap provides evidence of the areas that are most likely to be infected with *H. pylori*. Figure 2 shows the basis of the deep learning model making a correct decision (first row) and on what basis it made a misdiagnosis (second row) in the image of *H. pylori* infection. Figure 3 depicts the reason why the deep learning model determined the normal EGDS image as normal (first row) and as infected (second row) using heatmaps.

Performance measures. To evaluate the diagnostic capacity of the DCNN models, quantitative measures of the overall classification accuracy, sensitivity, specificity, positive

© 2023 The Authors. *JGH Open published by Journal of Gastroenterology and Hepatology* Foundation and John Wiley & Sons Australia, Ltd.

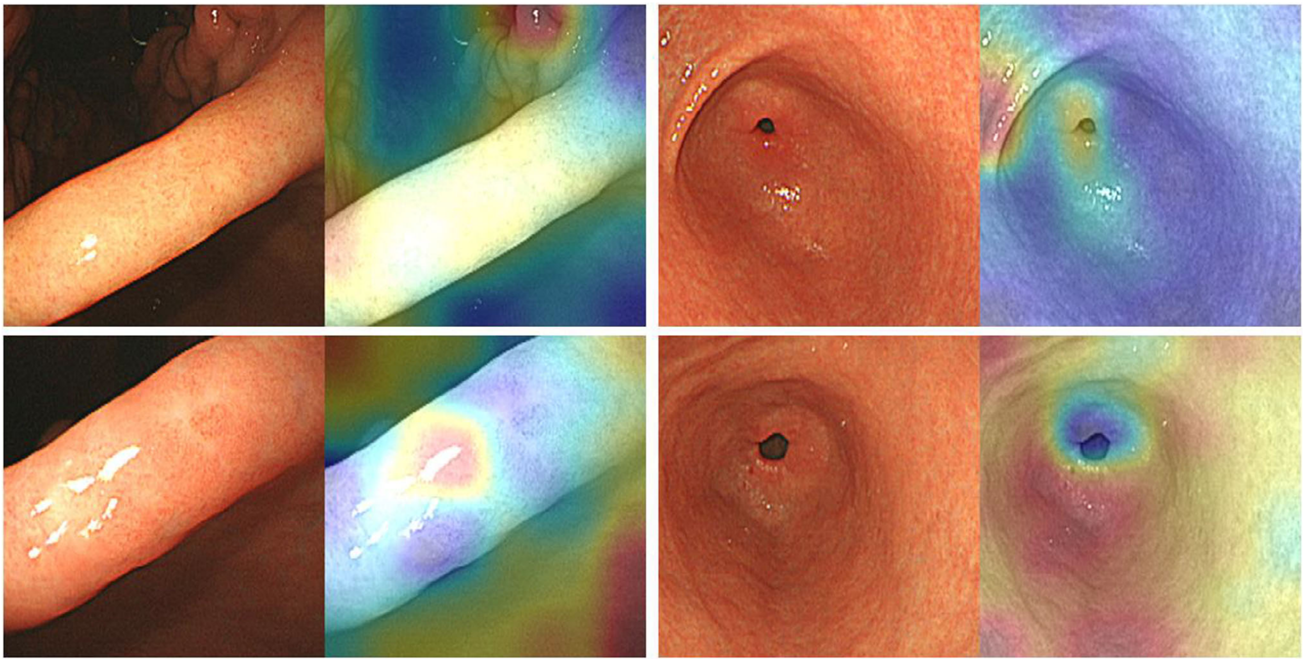

Figure 2 Gradient-weighted class activation map (Grad-CAM) result of overlaying the heatmap on infected images. The first and third columns show *Helicobacter pylori*-infected images. The second and fourth columns show the Grad-CAM results, which overlie the heatmap on the first and third column images, respectively. Grad-CAM results were generated using the InceptionResnet-v2 model. The first-row images are the true positive images, and the second-row images are the false negative images. Red areas show highly activated regions, and blue areas depict less activated regions.

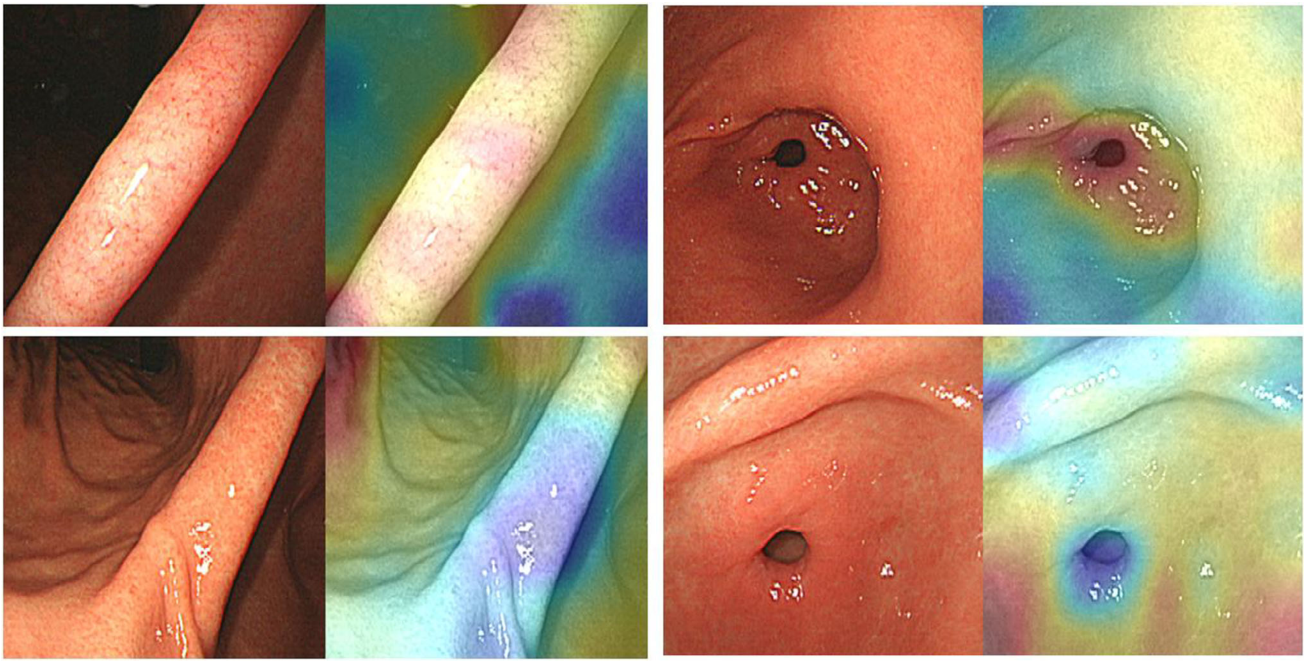

Figure 3 Gradient-weighted class activation map (Grad-CAM) result of overlaying the heatmap on uninfected images. The first and third columns represent the uninfected images. The second and fourth columns show the Grad-CAM results, which overlie the heatmap on the first and third column images, respectively. The first-row images are true negative images, and the second-row images are false positive images. Grad-CAM results were generated using the InceptionResnet-v2 model. Red areas show highly activated regions, and blue areas depict less activated regions.

© 2023 The Authors. *JGH Open published by Journal of Gastroenterology and Hepatology* Foundation and John Wiley & Sons Australia, Ltd.

predictive value, and negative predictive value were calculated as follows:

$$\text{Accuracy} = \frac{\text{True positive} + \text{true negative}}{\text{True positive} + \text{false positive} + \text{false negative} + \text{true negative}} \times 100\%,$$

$$\text{Sensitivity} = \frac{\text{True positive}}{\text{True positive} + \text{false negative}} \times 100\%,$$

$$\text{Specificity} = \frac{\text{True negative}}{\text{False positive} + \text{true negative}} \times 100\%,$$

$$\text{Positive predictive value} = \frac{\text{True positive}}{\text{True posirive} + \text{false positive}} \times 100\%,$$

$$\text{Negative predictive value} = \frac{\text{True negative}}{\text{False negative} + \text{true negative}} \times 100\%.$$

AUC is the area under the receiver operating characteristic (ROC) curve for the computed values of (1 – specificity) and sensitivity. The best cutoff point for each DCNN model to discriminate *H. pylori* infection was determined using the maximal Youden index (sensitivity + specificity – 1).[26] As the number of EGDS images in some regions was small, we grouped the cardia and fundus, lower body GC and lower body LC, and upper body GC and upper body LC into one region. This grouping was designed based on the anatomical regions of the stomach and the endoscopic images that were frequently displayed together.

To compare the six AUCs, we applied the nonparametric DeLong test.[27] Once the overall *P*-values indicated a significant difference, pairwise comparisons were made to calculate the *P*-value for each pair of models. The generalized estimating equation (GEE) method was used to compare sensitivity, specificity, accuracy, positive predictive value, and negative predictive value between the DCNN and ensemble models.[28] All analyses were performed using the SAS software (version 9.4; SAS Institute, Cary, NC, USA).

Results

Table 1 shows the diagnostic performances of the five DCNN models and the ensemble model when all EGDS images were used for training, validation, and test data. The best models for ResNet-101, Xception, Inception-v3, InceptionResnet-v2, and DenseNet-201 were obtained when the learning rates were 5e-5, 1e-4, and 5e-4, and L2 regularizations were 1e-3 and 5e-4, respectively. Among the models, the ensemble model, which was generated by averaging the output probabilities of the best models, achieved the best AUC (0.867), specificity (78.44%), accuracy (80.28%), positive predictive value (82.66%), and negative predictive value (77.37%). The highest sensitivity (83.75%) was obtained by the Inception-v3 model. A comparison of the ROC curves of the DCNN and the ensemble models is shown in Figure 4. Based on the DeLong test, the AUCs of the five DCNN models and the ensemble model were significantly different.

There were significant differences in sensitivity, specificity, accuracy, positive predictive value, and negative predictive value between the DCNN and ensemble models. The *P*-values listed in Table 1 are the overall *P*-values, which show the overall difference between each DCNN model, including the ensemble model.

Table 2 shows the AUC of the DCNN and ensemble models when EGDS images are trained for each sub-anatomical categories of stomach. For the sub-anatomical category, DenseNet-201 had the highest AUC (0.765) in the angle region, whereas the ensemble model achieved the best AUC for the other regions. The AUCs for the antrum, cardia and fundus, lower body GC and LC, and upper body GC and LC regions were 0.842, 0.826, 0.718, and 0.858, respectively, when the ensemble model was used. Compared with the results of dealing with all EGDS images, sub-anatomical category cases achieved a slightly lower AUC. Figure 4 depicts the ROC curves of different sub-anatomical regions with the ensemble model. The AUCs according to the sub-anatomical categories of the stomach were lower in the ensemble model than when the entire EGDS image was used.

Discussion

For the purpose of this study, it was assumed that the presence of *H. pylori* in any region of the stomach indicated an infection. Furthermore, it was hypothesized that *H. pylori* infection would result in variations in mucosal alterations depending upon the anatomical location within the stomach; this hypothesis was confirmed through the study's findings. In this study, we investigated the diagnostic performance of five DCNN models and an ensemble model to evaluate *H. pylori* infections using EGDS images. As shown in Table 1, all DCNN models achieved reasonable diagnostic performance, but the ensemble model showed better diagnostic performance than the DCNN models. Five DCNN models had AUCs of 0.83 or higher, while the ensemble model had the highest AUC of 0.867 when all EGDS images were used for training, validation, and test data. These findings suggest that the DCNN can be used to diagnose *H. pylori* infection.

Four DCNN models and the ensemble model had the highest AUC (0.858) in upper body GC and LC when images of each sub-anatomical category of the stomach were used. The reason for the higher AUC in the upper body GC and LC compared to other regions appears to be that mucosal changes were easier to observe compared to other regions; the number of pictures used for training was also a factor. Among the sub-anatomical categories of the stomach, the ensemble model showed the best AUCs for all except the angular region. In the angular region, DenseNet-201 had the best AUC (0.765). In a previous study, biopsy results revealed that the antrum had more *H. pylori*-associated gastritis than other regions.[29] In another study, the antrum and angular incisure showed more severe inflammation than the gastric body.[30] The AUC of the antrum was 0.842 in this study, which was the second highest after the upper body GC and LC. Previous reports have indicated that the antrum has greater mucosal changes due to *H. pylori*-associated gastritis because its AUC is higher than that of the other regions. However, the amount of training data for sub-anatomical categories that came out lower than the upper body GC and LC is different, which may be attributed to insufficient training. As shown in Tables 1 and 2, the findings of the sub-anatomical analysis appear to be slightly inferior to the results of processing all EGDS images in the diagnosis of *H. pylori*. This is most likely because the size of the data is not sufficient for training despite using a pretrained model. On an experimental basis, the AUC

© 2023 The Authors. *JGH Open published by Journal of Gastroenterology and Hepatology* Foundation and John Wiley & Sons Australia, Ltd.

Table 1 Diagnostic performance among deep convolutional neural network models when all esophagogastroduodenoscopy images are used

DCNN models	AUC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)	Positive predictive value (95% CI)	Negative predictive value (95% CI)
ResNet-101	0.830 (0.815–0.846)	75.07 (72.84–77.30)	76.97 (74.54–79.40)	75.91 (74.27–77.55)	80.38 (78.27–82.49)	71.06 (68.55–73.57)
Xception	0.833 (0.817–0.849)	77.20 (75.04–79.36)	75.58 (73.10–78.06)	76.49 (74.86–78.12)	79.90 (77.80–82.00)	72.51 (69.99–75.03)
Inception-v3	0.834 (0.818–0.850)	83.75 (81.85–85.65)	69.61 (66.96–72.26)	77.48 (75.88–79.08)	77.60 (75.54–79.66)	77.31 (74.76–79.86)
Inception Resnet-v2	0.847 (0.831–0.862)	83.26 (81.34–85.18)	71.17 (68.56–73.78)	77.91 (76.32–79.50)	78.40 (76.35–80.45)	77.18 (74.66–79.70)
DenseNet-201	0.846 (0.830–0.861)	79.41 (77.33–81.49)	75.41 (72.93–77.89)	77.64 (76.04–79.24)	80.24 (78.18–82.30)	74.44 (71.94–76.94)
Ensemble	0.867 (0.853–0.881)	81.75 (79.76–83.74)	78.44 (76.07–80.81)	80.28 (78.75–81.81)	82.66 (80.70–84.62)	77.37 (74.97–79.77)
P-value	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

The *P*-value represents the overall difference between each DCNN model including the ensemble model.

AUC, area under the curve; CI, confidence interval; DCNN, deep convolutional neural network.

Table 2 Area under the curve among deep convolutional neural network models when esophagogastroduodenoscopy images are trained for each sub-anatomical categories of stomach

DCNN models (95% CI)	Angle (n = 960)	Antrum (n = 2914)	Cardia and fundus (n = 1343)	Lower body GC and LC (n = 568)	Upper body GC and LC (n = 3110)
ResNet-101	0.693 (0.632–0.755)	0.810 (0.781–0.839)	0.801 (0.758–0.844)	0.701 (0.617–0.784)	0.829 (0.803–0.856)
Xception	0.656 (0.592–0.720)	0.801 (0.771–0.831)	0.800 (0.756–0.844)	0.695 (0.613–0.777)	0.837 (0.811–0.864)
Inception-v3	0.664 (0.600–0.729)	0.817 (0.788–0.846)	0.747 (0.699–0.794)	0.641 (0.557–0.724)	0.836 (0.810–0.862)
Inception Resnet-v2	0.739 (0.680–0.799)	0.811 (0.782–0.839)	0.820 (0.778–0.861)	0.711 (0.630–0.793)	0.836 (0.810–0.862)
DenseNet-201	0.765 (0.707–0.823)	0.831 (0.803–0.859)	0.748 (0.700–0.795)	0.709 (0.627–0.791)	0.810 (0.782–0.838)
Ensemble	0.753 (0.694–0.812)	0.842 (0.815–0.869)	0.826 (0.786–0.867)	0.718 (0.640–0.797)	0.858 (0.833–0.883)
P-value	<0.0001	<0.0001	<0.0001	0.0109	<0.0001

The *P*-value represents the overall difference between each DCNN model including the ensemble model.

CI, confidence interval; DCNN, deep convolutional neural network; GC, greater curvature; LC, lesser curvature.

© 2023 The Authors. *JGH Open published by Journal of Gastroenterology and Hepatology* Foundation and John Wiley & Sons Australia, Ltd.

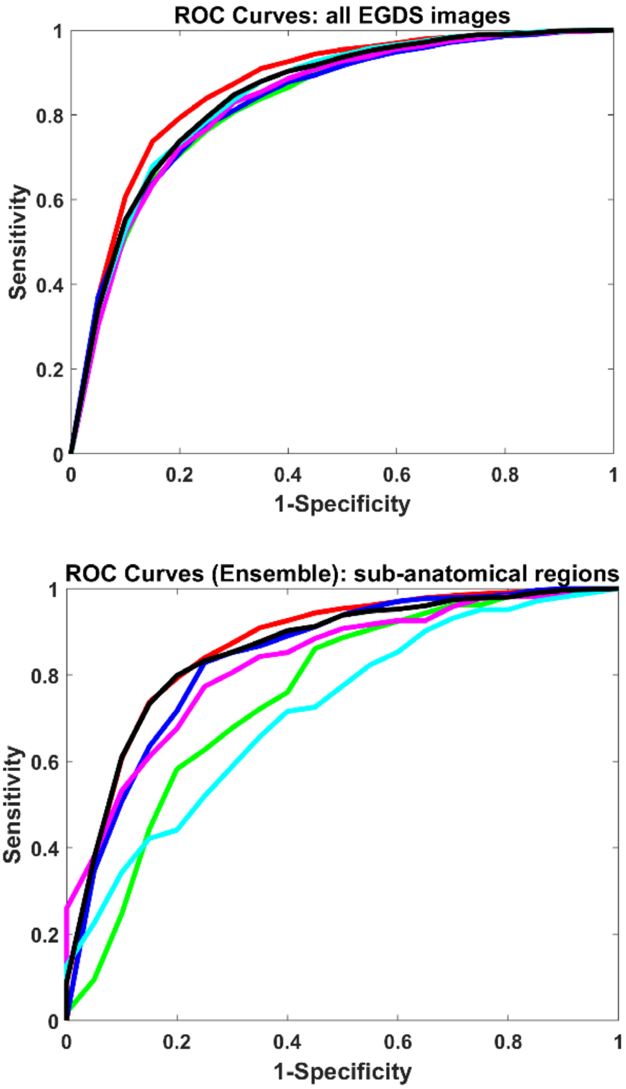

Figure 4 Receiver operating characteristic curves of deep convolutional neural network and ensemble models. Comparison of receiver operating characteristic (ROC) curves among deep convolutional neural network (DCNN) and ensemble models when all esophagogastroduodenoscopy (EGDS) images were used (upper panel). Comparison of ROC curves of the ensemble model when each sub-anatomical category EGDS image is used (lower panel). ——, ensemble (RUC = 00867); ——, ResNet-101 (AUC = 0.830); ——, Xception (AUC = 0.8333); ——, Inception-v3 (AUC = 0.834); ——, InceptionResnet-v2 (AUC = 0.847); ——, DenseNet-201 (AUC = 0.846); ——, All (AUC = 0.867); ——, Angle (AUC = 0.753); ——, Antrum (AUC = 0.842); ——, cardian and fundus (AUC = 0.826); ——, lower body GC and LC (AUC = 0.718); ——, upper body GC and LC (AUC = 0.858).

values of the angular region (960 training images) and lower body GC and LC regions (568 training images) were 0.753 and 0.718, respectively, which were smaller than the AUC values of the other regions. Because the number of pictures taken in this area was small, and the image quality was poor, the angle and lower body GC and LC were small, making it difficult to use for training. With regard to the angle component, it was determined that utilizing it for training posed a challenge due to the frequent variations in the shape and angle of the images captured by the endoscopist. The underlying reason seems to be that the number of training images used in these two domains was too small for adequate training. The AUCs of the antrum region (3914 training images), cardia and fundus regions (1343 training images), and upper-body GC and LC regions (3110 training images) were 0.842, 0.826, and 0.858, respectively, which approximated the values obtained when all EGDS images were used.

In a previous DCNN study using images of a large number of patients, the sensitivity, specificity, and accuracy of *H. pylori* infection diagnosis were 80% or higher, and the AUC was 0.89 or higher, although this was a small-scale study.[12] These findings were consistent with those in another study reporting an AUC of 0.956 and a sensitivity and specificity of 86.7%.[31] In a study using the LCI-CAD (linked color imaging-computer-aided detection) system, accuracy was 84.2% without infection, 82.5% with current infection, and 79.2% after disinfection treatment.[32] A previous meta-analysis had shown that AI was a reliable tool for the endoscopic diagnosis of *H. pylori* infection.[33] For the prediction of *H. pylori* infection, the pooled sensitivity, specificity, and AUC of AI were 0.87, 0.86, and 0.92, respectively. However, in a meta-analysis including eight studies, it was not possible to compare the diagnostic usefulness by applying the machine learning method to each part of the stomach or to evaluate the difference in diagnostic usefulness according to the machine learning methods. These results show similar or higher sensitivity, specificity, and AUC as those of the present study, which is likely due to the fact that this study focused on cases in which *H. pylori* infection was strongly suspected. AUC measures how well a model can distinguish between classes. The higher the AUC, the better the model can differentiate between patients with and without infection.[34]

For endoscopic diagnosis of *H. pylori* infection, no diagnostic technique has proven to be accurate.[11] Several studies have examined the link between endoscopic findings and *H. pylori* infection.[35-37] Changes in the mucous membranes were visually characterized as *H. pylori*-infected gastritis. Mucosal alterations, such as regular arrangement of collecting venules (RAC), atrophy, intestinal metaplasia, larger folds, nodularity, diffuse redness, and others, were used as diagnostic criteria for *H. pylori* infection in the Kyoto classification released in 2013.[38] Because of possible intra- or inter-observer variability in the optical diagnosis of *H. pylori*-infected mucosa, these endoscopic alterations are not objective indicators.[13] The majority of the CLO tests were performed, as in this study, by endoscopists and involved cases in which *H. pylori* infection was suspected on visual inspection of the gastric mucosa. This could explain why the diagnostic rate in this study was slightly lower than those reported in the previous studies.

The findings of this study suggest that even when *H. pylori* infection is suspected, AI can be used to screen infection. The strength of this study is that images were captured using endoscopic instruments by the same manufacturer and were classified by a single specialist, ensuring consistent image quality. A further strength of this study is that it compared sub-anatomical regions of the stomach and used various deep learning methods to evaluate the most appropriate

© 2023 The Authors. *JGH Open published by Journal of Gastroenterology and Hepatology* Foundation and John Wiley & Sons Australia, Ltd.

DCNN method and the optimal sub-anatomical region for AI application in the diagnosis *H. pylori* infection. Furthermore, despite the use of visual images of mucous membranes with suspected *H. pylori* infection, image analysis yielded relatively high diagnostic power.

Some limitations of this study should be noted. First, both development and test datasets were gathered from a single hospital. Validation with images from other facilities and various endoscopic devices and procedures may improve the generalizability of our findings. However, we used over 10 000 images in this investigation, so this constraint may be resolved. Second, the CLO test, which is used to confirm *H. pylori* infection status, is simple, efficient, and cost effective and has a high degree of sensitivity (around 90%) and specificity (between 95% and 100%) in identifying the presence of *H. pylori*. This may have affected our evaluation of the diagnostic ability of DCNN. This issue could be addressed by including information about the technique of validating *H. pylori* infection status in the design of DCNN. In order to increase the diagnostic rate, samples were obtained from two separate locations within both the antrum and the body of the stomach; however, in most cases, it has the drawback of being confirmed by a single EGDS test. Third, the CLO test was performed only on patients with suspected *H. pylori* infection; the study compared only the negative and positive groups, and there was no healthy control group. As a result, the diagnosis rate may be lower than reported in studies involving healthy controls.

This study revealed that endoscopic images can detect *H. pylori* infection when using a deep learning method. It is envisaged that applying this method to data from large-scale national health check-ups will help lessen the disparity in *H. pylori* infection evaluations between EGDS observers and more properly assess people with suspected *H. pylori* infection. As a result, this will reduce medical costs related to CLO testing.

Conclusion

Our findings show that the DCNN built to perform automatic analysis of stored images could help with accurate *H. pylori* infection screening and identify patients who require confirmation tests.

Acknowledgments

This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HR21C1003).

References

1 Hunt R, Xiao SD, Megraud F *et al*. *Helicobacter pylori* in developing countries. World gastroenterology organisation global guideline. *J. Gastrointest. Liver Dis.* 2011; **20**: 299–304.

2 Lee JH, Ahn JY, Choi KD *et al.* Nationwide antibiotic resistance mapping of *Helicobacter pylori* in Korea: a prospective multicenter study. *Helicobacter*. 2019; **24**: e12592.

3 Chey WD, Leontiadis GI, Howden CW, Moss SF. ACG clinical guideline: treatment of *Helicobacter pylori* infection. *Am. J. Gastroenterol.* 2017; **112**: 212–39.

4 Bang CS, Lee JJ, Baik GH. The most influential articles in *Helicobacter pylori* research: a bibliometric analysis. *Helicobacter*. 2019; **24**: e12589.

5 Correa P. A human model of gastric carcinogenesis. *Cancer Res.* 1988; **48**: 3554–60.

6 Zhang C, Yamada N, Wu YL, Wen M, Matsuhisa T, Matsukura N. *Helicobacter pylori* infection, glandular atrophy and intestinal metaplasia in superficial gastritis, gastric erosion, erosive gastritis, gastric ulcer and early gastric cancer. *World J Gastroenterol*. 2005; **11**: 791–6.

7 Sung H, Ferlay J, Siegel RL *et al*. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 2021; **71**: 209–49.

8 IARC working group on the evaluation of carcinogenic risks to humans, et al. Infection with *Helicobacter pylori*, in *Schistosomes*, liver flukes and *Helicobacter pylori*. International Agency for Research on Cancer, 1994 .

9 Plummer M, Franceschi S, Vignat J, Forman D, de Martel C. Global burden of gastric cancer attributable to *Helicobacter pylori*. *Int. J. Cancer*. 2015; **136**: 487–90.

10 Fock KM, Katelaris P, Sugano K *et al*. Second Asia–Pacific consensus guidelines for *Helicobacter pylori* infection. *J. Gastroenterol. Hepatol.* 2009; **24**: 1587–600.

11 Glover B, Teare J, Patel N. A systematic review of the role of non-magnified endoscopy for the assessment of *H. pylori* infection. *Endos. Intern. Open*. 2020; **8**: E105–14.

12 Shichijo S, Endo Y, Aoyama K *et al*. Application of convolutional neural networks for evaluating *Helicobacter pylori* infection status on the basis of endoscopic images. *Scand. J. Gastroenterol.* 2019; **54**: 158–63.

13 Yasuda T, Hiroyasu T, Hiwa S *et al*. Potential of automatic diagnosis system with linked color imaging for diagnosis of *Helicobacter pylori* infection. *Dig. Endosc.* 2020; **32**: 373–81.

14 Mccarthy JF, Marx KA, Hoffman PE *et al*. Applications of machine learning and high-dimensional visualization in cancer detection, diagnosis, and management. *Ann. N. Y. Acad. Sci.* 2004; **1020**: 239–62.

15 LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015; **521**: 436–44.

16 He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. Proceedings of the IEEE international conference on computer vision; 2015.

17 Cho BJ, Bang CS. Artificial intelligence for the determination of a management strategy for diminutive colorectal polyps: hype, hope, or help. *LWW*. 2020: 70–2.

18 Cho B-J, Bang CS, Park SW *et al*. Automated classification of gastric neoplasms in endoscopic images using a convolutional neural network. *Endoscopy*. 2019; **115**: 1121–9.

19 Moon SW, Kim TH, Kim HS *et al*. United rapid urease test is superior than separate test in detecting *Helicobacter pylori* at the gastric antrum and body specimens. *Clinic. Endos.* 2012; **45**: 392–6.

20 He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition; 2016.

21 Chollet F. Xception: Deep learning with depthwise separable convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition; 2017.

22 Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. Proceedings of the IEEE conference on computer vision and pattern recognition; 2016.

23 Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, inception-resnet and the impact of residual connections on learning. Thirty-first AAAI conference on artificial intelligence; 2017.

© 2023 The Authors. *JGH Open published by Journal of Gastroenterology and Hepatology* Foundation and John Wiley & Sons Australia, Ltd.

24 Huang G, Liu Z, Maaten Laurens Van Der, Weinberger KQ. Densely connected convolutional networks. Proceedings of the IEEE conference on computer vision and pattern recognition; 2017.

25 Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: Visual explanations from deep networks via gradient-based localization. Proceedings of the IEEE international conference on computer vision; 2017.

26 Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950; **3**: 32–5.

27 DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988; **44**: 837–45.

28 Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986; **73**: 13–22.

29 Trindade LMDF, Menezes LBO, Souza Neta AM *et al*. Prevalence of *Helicobacter pylori* infection in samples of gastric biopsies. *Gastroenterology Res.* 2017; **10**: 33–41.

30 Álvares MMD, Marino M, Oliveira CA *et al*. Características da gastrite crônica associada a *Helicobacter pylori*: aspectos topográficos, doenças associadas e correlação com o status cagA. *Jornal Brasileiro de Patologia e Medicina Laboratorial*. 2006; **42**: 51–9.

31 Itoh T, Kawahira H, Nakashima H, Yata N. Deep learning analyzes *Helicobacter pylori* infection by upper gastrointestinal endoscopy images. *Endosc Intern Open*. 2018; **6**: E139–44.

32 Nakashima H, Kawahira H, Kawachi H, Sakaki N. Endoscopic three-categorical diagnosis of *Helicobacter pylori* infection using linked color imaging and deep learning: a single-center prospective study (with video). *Gastric Cancer*. 2020; **23**: 1033–40.

33 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**: e21983.

34 Narkhede S. Understanding auc-roc curve. *Towards Data Sci.* 2018; **26**: 220–7.

35 Watanabe K, Nagata N, Nakashima R *et al*. Predictive findings for *Helicobacter pylori*-uninfected,-infected and-eradicated gastric mucosa: validation study. *World J. Gastroenterol.* 2013; **19**: 4374–9.

36 Olmez S, Aslan M, Erten R, Sayar S, Bayram I. The prevalence of gastric intestinal metaplasia and distribution of *Helicobacter pylori* infection, atrophy, dysplasia, and cancer in its subtypes. *Gastroenterol. Res. Pract.* 2015; **2015**: 1–6.

37 Laine L, Cohen H, Sloane R, Marin-Sorensen M, Weinstein WM. Interobserver agreement and predictive value of endoscopic findings for *H. pylori* and gastritis in normal volunteers. *Gastrointest. Endosc.* 1995; **42**: 420–3.

38 Toyoshima O, Nishizawa T, Koike K. Endoscopic Kyoto classification of *Helicobacter pylori* infection and gastric cancer risk diagnosis. *World J. Gastroenterol.* 2020; **26**: 466–77.

© 2023 The Authors. *JGH Open published by Journal of Gastroenterology and Hepatology* Foundation and John Wiley & Sons Australia, Ltd.