Title:

Detection of gastritis by a deep convolutional neural network from double-contrast upper gastrointestinal barium X-ray radiography

Short title:

Gastritis detection by deep learning

Authors and Institution:

Ren Togo[1], Nobutake Yamamichi[2], Katsuhiro Mabe[3], Yu Takahashi[2], Chihiro Takeuchi[2], Mototsugu Kato[3], Naoya Sakamoto[4], Kenta Ishihara[1], Takahiro Ogawa[1], Miki Haseyama[1]

1. Graduate School of Information Science and Technology, Hokkaido University, N-14, W-9, Kita-ku, Sapporo, Hokkaido, 060-0814, Japan. E-mail: togo@lmd.ist.hokudai.ac.jp

2. Department of Gastroenterology, Graduate School of Medicine, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-8655, Japan.

3. Department of Gastroenterology, National Hospital Organization Hakodate Hospital, 18-16, Kawahara-cho, Hakodate City, Hokkaido 041-8512, Japan

4. Department of Gastroenterology, Hokkaido University Graduate School of Medicine, Sapporo 060-8648, Japan

Corresponding author: Katsuhiro Mabe, Department of Gastroenterology, National Hospital Organization Hakodate Hospital, Hakodate, Hokkaido, 041-8512, Japan.

TEL: +81-0138-51-6281

FAX: +81-0138-51-6288

E-mail: katsumabe@me.com

Word count:

4553

Abbreviation lists:

DCNN: Deep convolutional neural network

Ha: Harmonic mean

H. pylori: *Helicobacter pylori*

PG: Pepsinogen

ROC curve: Receiver operating characteristic curve

ROI: Region of interest

Se: Sensitivity

Sp: Specificity

UGI-ES: Upper gastrointestinal endoscopy

UGI-XR: Double-contrast upper gastrointestinal barium X-ray radiography

ABSTRACT

Background: Deep learning has become a new trend of image recognition tasks in the field of medicine. We developed an automated gastritis detection system using double-contrast upper gastrointestinal barium X-ray radiography.

Methods: A total of 6,520 gastric X-ray images obtained from 815 subjects were analyzed. We designed a deep convolutional neural network (DCNN)-based gastritis detection scheme and evaluated the effectiveness of our method. The detection performance of our method was compared with that of ABC (D) stratification.

Results: Sensitivity, specificity, and harmonic mean of sensitivity and specificity of our method were 0.962, 0.983, and 0.972, respectively, and those of ABC (D) stratification were 0.925, 0.998, and 0.960, respectively. Although there were 18 false negative cases in ABC (D) stratification, 14 of those 18 cases were correctly classified into the positive group by our method.

Conclusions: Deep learning techniques may be effective for evaluation of gastritis/non-gastritis. Collaborative use of DCNN-based gastritis detection systems and ABC (D) stratification will provide more reliable gastric cancer risk information.

Keywords:

Deep convolutional neural network▪Artificial intelligence▪Gastritis▪Double-contrast upper gastrointestinal barium X-ray radiography

1. INTRODUCTION

Gastric cancer has remained a burdensome disease in East Asian countries with high mortality rates [1,2]. It has been revealed that gastritis is mainly caused by *Helicobacter pylori* (*H. pylori*) infection [3], and such a condition increases the risk of gastric cancer. Hence, evaluation of gastric cancer risk for primary prevention has become an urgent issue [4,5]. Several inspections have been used for evaluation of gastric cancer risk and detection of gastric cancer. ABC (D) stratification based on serum anti-*H. pylori* IgG and pepsinogens (PG I and PG II) has shown to be effective for the evaluation of gastric cancer risk [6]. However, false negative cases in ABC (D) stratification often include high gastric cancer risk individuals [7]. On the other hand, double-contrast upper gastrointestinal barium X-ray radiography (UGI-XR) [8] is the conventional gold standards for detection of gastric cancer as a mass screening program in Japan. Nowadays, it is expected that UGI-XR will enable not only the detection of gastric cancer but also the evaluation of *H. pylori* infection status and atrophic changes of the stomach for realizing the risk-based mass screening. However, in addition to detection of gastric cancer, evaluation of the condition of gastric folds or mucosal surfaces increases the image interpreting time. Computer-aided diagnosis systems that analyze gastric X-ray images and provide supporting information for medical specialists are desired.

In the field of medical image analysis, deep convolutional neural networks (DCNNs) [9] have recently become a new trend of machine learning techniques since their recognition performance has exceeded that of other conventional machine learning techniques [10]. DCNNs enable extraction of clinically important features by a backpropagation algorithm, and these extracted features are useful for improving the recognition performance compared to handcrafted features that are designed by heuristic knowledge [11].

In this study, we examined the effectiveness of DCNNs for detection of gastritis using UGI-XR. We developed a DCNN-based automated gastritis detection method and compared its detection performance with that of ABC (D) stratification. Experimental results indicated that DCNNs can evaluate the condition of the gastric mucosa. The results of this pilot study will contribute to the acceleration of risk-based gastric cancer screening.

2. MATERIALS AND METHODS

We developed an automated DCNN-based gastritis detection method and we evaluated its effectiveness retrospectively. This study was reviewed and approved by the institutional review board. Patients were not required to give informed consent to the study because the analysis used anonymous data that were obtained after each patient agreed to inspections by written consent.

Study subjects

In this retrospective study, 815 subjects who underwent UGI-XR, upper gastrointestinal endoscopy (UGI-ES), and ABC (D) stratification at The University of Tokyo Hospital in 2010 were reviewed [12]. The gold standard of our DCNN-based gastritis detection method was the diagnostic results of UGI-XR and UGI-ES. The values of PG and anti-*H. pylori* titers were measured for ABC (D) stratification. Serum anti-*H. pylori* IgG and pepsinogens were measured by commercial kits (E Plate Eiken H. pylori, Eiken Chemical Co., Ltd., Tokyo, Japan). In ABC (D) stratification, the values of PG and anti-*H. pylori* titers are combined for evaluation of the risk of gastric cancer. As shown in Table 1, group A is defined as a very low gastric cancer risk group, group B is defined as a middle-risk group, and groups C and D are defined as high-risk groups. It should be noted that group D is generally included in group C [13]. We compared the

detection performance of our method and ABC (D) stratification.

The subject's selection flowchart is shown in Fig. 1. The UGI-XR-based evaluation was conducted by the established four-grade atrophic level classification, namely, normal, mild, moderate, and severe [14]. Subjects who had diagnostic results of "normal" in UGI-XR were allocated to a negative group, and the other subjects were allocated to a positive group. Also, UGI-ES-based evaluation was conducted by the Kimura-Takemoto seven-grade classification, namely, no atrophic change (C0), three closed types of atrophic gastritis (C1, C2, C3), and three open types of atrophic gastritis (O1, O2, O3) [15,16]. Subjects who had diagnostic results of "C0" in UGI-ES were allocated to a negative group, and subjects diagnosed as "C2-O3" were allocated to a positive group. Subjects diagnosed as "C1" were excluded from our study subjects since this class is defined as the atrophic borderline in the Kimura-Takemoto seven-grade classification. Data for subjects who had the same diagnostic results for UGI-XR and UGI-ES were analyzed in this study. In ABC (D) stratification, subjects in group A were allocated to a negative group and subjects in groups B and C (D) were allocated to a positive group in this study.

In UGI-XR, multiple gastric X-ray images were taken for evaluation of the condition of the stomach. We used gastric X-ray images obtained at the following eight positions: 1) double-contrast frontal view of the stomach in the supine position, 2) double-contrast right anterior oblique view of the stomach in the near-supine position, 3) double-contrast left anterior oblique view of the stomach in the near-supine position, 4) double-contrast frontal view of the stomach in the prone position with the head down, 5) double-contrast frontal view of the stomach in the prone standing position, 6) double-contrast left lateral view of the stomach in the horizontal position, 7) double-contrast left anterior oblique view of the stomach in the near-supine position and 8) double-contrast right anterior oblique view of the stomach in the near-supine half-standing position. All gastric X-ray images were 8-bit gray-scale and 2,048 × 2,048 pixels. Since gastric X-ray images were obtained in eight positions for each subject, 815 × 8 (6,520)

gastric X-ray images were used for training and testing procedures. We calculated the detection performance of our method and that of ABC (D) stratification based on the gold standard.

DCNN-based gastritis detection method

Artificial intelligence (AI) is the term first coined in 1956, and AI involves machines that can perform tasks that are characteristics of human intelligence. Machine learning is one of the effective approaches for achieving AI and is defined as the algorithms that can learn from large amounts of data without relying on rule-based programming. Among machine learning techniques, deep learning algorithms that mimic the biological structure of the brain have recently become a new trend in the field of image recognition. Particularly, DCNNs are one of the popular methods in deep learning techniques and are state-of-the-art methods in medical image recognition tasks. Typical DCNNs have several convolution and pooling layers, and their network architecture differs according to the target tasks or the amount of data. However, it is difficult to analyze gastric X-ray images used in this study compared to CT or MRI images due to the following problems.

Problem 1. Gastric X-ray images have high resolutions that have an effect on a parameter calculation cost of DCNNs.

Problem 2. Only an image-level annotation (gastritis or non-gastritis) result is provided since region-level annotation requires a great deal of labor.

We designed a DCNN architecture for solving the above problems. Figure 2 shows our designed DCNN-based gastritis detection scheme. The gastritis detection scheme is composed of the following two steps: (i) patch-based training and (ii) calculation of gastritis-related features. Each step uses different DCNNs: a patch extraction DCNN and a refined DCNN. The patch extraction DCNN is used to automatically find important regions related to gastritis from all regions of gastric X-ray images such as the region of interest (ROI) setting. The refined

DCNN is used to learn the differences in gastritis/non-gastritis and produce image features that can clearly represent gastritis characteristics. Extracted DCNN-based features are used for a support vector machine [17] classification.

First, in step (i), we divide gastric X-ray images into multiple patches and evaluate the risk of gastric cancer for solving Problems 1 and 2. All of the patches extracted from gastric X-ray images are used for training, and the patch extraction DCNN calculates relevance scores between each patch and image-level annotations. Then it can be considered that relevance scores of patches from regions related to gastritis/non-gastritis become high and those of patches from regions irrelevant to gastritis/non-gastritis become low. Since the histogram of relevance scores has a bimodal distribution, patches related to gastritis are automatically extracted from all patches by the Otsu thresholding method [18]. Extracted patches are used as new training data for step (ii).

In step (ii), we introduce a refined DCNN that is trained with only extracted patches in step (i). The refined DCNN learns the characteristics of gastritis/non-gastritis. Since pre-processed training data do not include patches irrelevant to gastritis/non-gastritis owing to the procedure of step (i), the refined DCNN can learn the differences in them effectively. However, there still remain several problems. Symptoms associated with gastric cancer risk are described in various regions of the stomach. In addition, the size of the stomach is different for each subject. Therefore, we consider such individual characteristics by applying bag-of-features representation [19]. Then we use the refined DCNN as a feature extractor and calculate new clinically important image features. Finally, each patch is evaluated by a support vector machine-based classifier for detection of gastritis/non-gastritis. For evaluation of test data, estimated results of each positions' gastric X-ray image were integrated by a simplest majority voting method.

Our system was developed on a Linux operating system (Ubuntu 14.04; Canonical, London, England) with the Caffe framework [20] and a single NVIDIA Tesla 80K GPU. Patch

making procedure was written in MATLAB codes and other procedures were written in python codes. For DCNN parameter settings, we determined the size of patches and its overlapping to 200 × 200 pixels and 50 pixels, respectively. Consequently, a total of 3,542,664 patches were used for training our DCNNs. The batch size was set to 128, and the number of epochs was set to 50 for all experiments.

Statistical analysis

The verification method was 5-fold cross-validation. For evaluation of the detection performance of our method, sensitivity (Se = true positive / (true positive + false negative)), specificity (Sp = true negative / (true negative + false positive)), and harmonic mean of sensitivity and specificity (Ha = (2 × Se × Sp) / (Se + Sp)) were used. Note that Ha represents the total performance of gastritis detection. A receiver operating characteristic (ROC) curve was generated on the basis of detection results. The ROC curve was obtained by changing the threshold that determines gastritis/non-gastritis.

3. RESULTS

Our DCNN-based method achieved a performance equivalent to ABC (D) stratification.

Detection performance of our method and that of ABC (D) stratification are shown in Table 2. Sensitivity (gastritis), specificity (non-gastritis) and harmonic mean of our method were 0.962, 0.983, and 0.972, respectively, and those of ABC (D) stratification were 0.925, 0.998, and 0.960, respectively. The ROC curve of our method that was obtained by changing the threshold value for determining gastritis is shown in Fig. 3. Our DCNN showed high detection performance nearly equivalent to the performance of ABC (D) stratification when the

gold standard was set by image interpretation results of UGI-XR and UGI-ES.

Our DCNN-based method focused on clinically important regions for estimating gastritis/non-gastritis.

Figure 4 shows examples of original images and relevance scores of our method. The heat map of relevance scores is overlaid on the original images. The closer to yellow color the heat map becomes, the stronger is the relation of gastritis/non-gastritis, whereas blue color regions are irrelevant to gastritis/non-gastritis in our method. It was shown that our DCNN focused on regions related to gastritis/non-gastritis when estimating gastritis, even if the results of ABC (D) stratification were different.

Characteristics of false positive/negative cases were different between our DCNN method and ABC (D) stratification.

Figure 5 shows examples of false negative and false positive gastric X-ray images in our method. We can see that it is difficult to evaluate the condition of the gastric mucosa from these gastric X-ray images due to overlapping of the barium contrast medium. These cases were classified into the correct group by ABC (D) stratification. Although there were 18 false negative cases in ABC (D) stratification, 14 of those 18 cases were correctly classified into the positive group by our DCNN-based method. Furthermore, there was only one false positive case in ABC (D) stratification, this case was correctly classified into the negative group by our DCNN-based method.

As an additional experiment, we evaluated subjects diagnosed as “C1” by the Kimura-Takemoto seven-grade classification of UGI-ES. Subjects diagnosed as “C1” were excluded for training and testing procedures of our DCNN method since this class is defined as the atrophic

borderline and multi-stage risk evaluation is a more challenging task than a simple two-class (positive/negative) classification task. There were 37 cases diagnosed as "C1" by UGI-ES and they were excluded from the above main experiment. All the "C1" cases were classified into the positive group by our DCNN-based method. On the other hand, their ABC (D) stratification results were as follows. Group A: 24 cases, group B: 12 cases, and group C: 1 case. Namely, 35.1% of "C1" cases were classified into the positive group by ABC (D) stratification. We confirmed that the judgment tendency of our method and ABC (D) stratification was different from each other. DCNN-based risk evaluation has a potential to deal with subjects who have a slight atrophy as the positive group like UGI-ES, whereas many cases are judged as the negative group in ABC (D) stratification.

4. DISCUSSION

Deep learning algorithms, particularly DCNNs, have rapidly become a first choice for medical image analysis since they can automatically learn high-level representation features, *e.g.*, disease presence/absence. The number of reports on deep learning algorithms for medical images has been rapidly increasing in recent years [21]. However, most of those studies were based on MRI, microscopy, and CT images, and most of the target diseases were diseases in the brain, lung, and abdomen [21]. On the other hand, there have been few reports on UGI-XR for gastritis detection. The following two main reasons might cause such a situation. The first one is the resolution of images. Typical gastric X-ray images have 1,024 × 1,024 or 2,048 × 2,048 pixels, while typical CT and MRI images have 256 × 256 pixels. The size of images affects the cost of calculating parameters of DCNNs. It is still difficult to analyze high-resolution images in the field of machine learning. The second reason is the quality of the images. Although the DCNN training procedure including irrelevant regions can cause deterioration of detection

performance, a manual ROI setting operation is labor-intensive. In UGI-XR, the quality of gastric X-ray images is influenced by body movement, while CT and MRI can be executed in a condition with no body movement. It is difficult to set appropriate ROI regions related to the stomach and gastritis automatically.

In this study, we developed an automated gastritis detection method that can solve the above problems. To the best of our knowledge, this report is the first report on the potential of a DCNN for application to gastric cancer mass screening using UGI-XR. We compared the detection performance of our method and that of ABC (D) stratification. The results showed that the detection performance of our method was almost equivalent to that of ABC (D) stratification. Moreover, we confirmed that our DCNN-based method focused on clinically important regions when estimating gastritis/non-gastritis. Considering the clinical application of automated gastritis detection systems for the evaluation of gastric cancer risk, the number of false negative cases is a controversial issue. When there is a hidden region in an X-ray image or too much bulge by a blowing agent, condition of gastric folds and mucosa are difficult to be described. This situation can cause the false positive/negative classification in our method. We confirmed that characteristics of false positive/negative cases were different between DCNN-based method and ABC (D) stratification. This experimental results indicate that DCNN-based risk evaluation can make up for the disadvantage of ABC (D) stratification. Collaborative use of DCNN-based risk information and ABC (D) stratification-based risk information may enable more accurate evaluation of gastric cancer risk.

The establishment of effective gastric cancer mass screening is an urgent issue for the reduction of gastric cancer mortality rates [22]. Specifically, it is important to narrow down individuals who need UGI-ES examination by evaluating atrophic levels of the stomach and *H. pylori* infection status. It is certain that ABC (D) stratification based on the blood examination is a simple, non-inversive and rather inexpensive screening method to evaluate gastric cancer risk. However, several problems arise in this screening method. First, patients with severe gastric

mucosal atrophy often denote negative results in serum anti-*H. pylori* IgG test, which is known as the “pseudo A group” in ABC (D) stratification. Secondly, ABC (D) stratification can never be applied to the evaluation of gastric cancer risk after *H. pylori* eradication therapy. On the other hand, since UGI-XR can effectively cover such cases, evaluation of gastritis by UGI-XR plays a significant role in gastric cancer mass screening [23]. UGI-XR is the current gold standard inspection for the mass screening, and evaluation methods have already been established. However, evaluation of fold distributions and mucosal surfaces from gastric X-ray images is time-consuming and often subjective. On the other hand, roles for AI based on machine learning techniques have gradually been recognized in the field of medical image analysis [24–26]. It has been expected that machine learning techniques for analysis of gastric X-ray images may remarkably improve the efficiency of gastric cancer mass screening.

There are several limitations. First, the data used in this study were obtained from a single medical facility. Considering the application of deep learning techniques for automated gastritis detection, further investigations using data for various facilities will enhance the reliability of deep learning techniques. Our previous studies using 16,800 X-ray images from other medical facilities has already shown the effectiveness of machine learning techniques [27–29], it can be considered that deep learning techniques for UGI-XR have bright prospects. Second, we excluded *H. pylori*-eradicated subjects from our analysis in order to fairly compare the performance of our method and that of ABC (D) stratification. Ideally, the advantage of image-based supporting systems is that they can provide gastric cancer risk information from subjects who have experience of *H. pylori* eradication therapy by analyzing gastric X-ray images. The realization of this risk evaluation will have a great clinical impact in the field of medicine. However, since there have been few reports in which the effectiveness of deep learning techniques for UGI-XR is discussed, we focused on the evaluation of deep learning techniques for UGI-XR in this study. The next step of this study for *H. pylori*-eradicated subjects is one of our future works. One more thing we should try to do in the next step is a

multi-stage classification approach. Although the task of this study was gastritis/non-gastritis detection, ABC (D) stratification and the Kimura-Takemoto seven-grade classification can evaluate multi-stage risk information. We had tried to classify the gastric cancer risk into the three-grade using other datasets so far [27], and reported the effectiveness of the machine learning approach. Hence, we will take on this multi-stage evaluation using DCNN-based method as the next step of this study.

It is concluded that deep learning techniques for UGI-XR may be effective for gastritis detection. Although further investigation is needed, the results of this preliminary study will contribute to the acceleration of risk-based gastric cancer screening.

ACKNOWLEDGMENT

The clinical data were acquired at The University of Tokyo Hospital in Japan. This study was partly supported by Global Station for Big Data and Cybersecurity, a project of Global Institution for Collaborative Research and Education at Hokkaido University JSPS KAKENHI Grant Number JP17H01744.

DISCLOSURE

Dr. N. Sakamoto received honoraria from Gilead Sciences, Bristol-Myers-Squibb, MSD, Otsuka Pharmaceuticals, AbbVie, Daiichi Sankyo and Abott. Also, Dr. N. Sakamoto received commercial research funding from Gilead Sciences, Bristol-Myers-Squibb, MSD, Otsuka Pharmaceuticals, AbbVie, Daiichi Sankyo, Shionogi, Astellas, Takeda, Sumitomo Dainippon Pharma, EA pharma and Eisai.

REFERENCES

1. F. B. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012: Globocan 2012. Int J Cancer. 2015;136:E359--E386.

2. Jung K-W, Won Y-J, Kong H-J, et al. Cancer statistics in Korea: Incidence, mortality, survival, and prevalence in 2012. Cancer Res Treat. 2015;47:127–41.

3. The EUROGAST Study Group. An international association between Helicobacter pylori infection and gastric cancer. Lancet. Elsevier; 1993;341:1359–63.

4. Shimoyama T, Aoki M, Sasaki Y, et al. ABC screening for gastric cancer is not applicable in a Japanese population with high prevalence of atrophic gastritis. Gastric Cancer. 2012;15:331–4.

5. Zhu R, Chen K, Zheng Y-Y, et al. Meta-analysis of the efficacy of probiotics in Helicobacter pylori eradication therapy. World J Gastroenterol. Baishideng Publishing Group Inc; 2014;20:18013–21.

6. Kudo T, Kakizaki S, Sohara N, et al. Analysis of ABC (D) stratification for screening patients with gastric cancer. World J Gastroenterol. Baishideng Publishing Group Inc; 2011;17:4793–8.

7. Miura K, Okada H, Kouno Y, et al. Actual status of involvement of Helicobacter pylori infection that developed gastric cancer from group A of ABC (D) stratification - study of early gastric cancer cases that underwent endoscopic submucosal dissection. Digestion. Karger Publishers; 2016;94:17–23.

8. Oshima A, Hirata N, Ubukata T, et al. Evaluation of a mass screening program for stomach cancer with a case‐control study design. Int J Cancer. Wiley Subscription Services, Inc., A Wiley Company; 1986;38:829–33.

9. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Adv Neural Inf Process Syst. 2012. p. 1–9.

10. Suzuki K. Overview of deep learning in medical imaging. Radiol Phys Technol. 2017;10:257–73.

11. Yann LeCun, Yoshua Bengio, Geoffrey Hinton. Deep learning. Nature. 2015;521:436–44.

12. Yamamichi N, Hirano C, Takahashi Y, et al. Comparative analysis of upper gastrointestinal endoscopy, double-contrast upper gastrointestinal barium X-ray radiography, and the titer of serum anti-Helicobacter pylori IgG focusing on the diagnosis of atrophic gastritis. Gastric Cancer. Springer Japan; 2016;19:670–5.

13. Itoh T, Saito M, Marugami N, et al. Correlation between the ABC classification and radiological findings for assessing gastric cancer risk. Jpn J Radiol. Springer Japan; 2015;33:636–44.

14. Dheer S, Levine MS, Redfern RO, et al. Radiographically diagnosed antral gastritis: Findings in patients with and without Helicobacter pylori infection. Br J Radiol. British Institute of Radiology; 2002;75:805–11.

15. Kimura K, Takemoto T. An endoscopic recognition of the atrophic border and its significance in chronic gastritis. Endoscopy. © Georg Thieme Verlag, Stuttgart; 1969;1:87–97.

16. Kimura K. Chronological transition of the fundic-pyloric border determined by stepwise biopsy of the lesser and greater curvatures of the stomach. Gastroenterology. 1972;63:584–92.

17. Cortes C. Support-vector networks. Mach Learn. Springer; 1995;20:273–97.

18. Otsu N. A threshold selection method from gray-level. IEEE Trans Syst Man Cybern. 1979;9:62–6.

19. Venegas-Barrera CS, Manjarrez J. Visual categorization with bags of keypoints. Work Stat Learn Comput vision, ECCV. 2004. p. 1–2.

20. Jia Y, Shelhamer E, Donahue J. Caffe: Convolutional architecture for fast feature embedding. Int Conf Multimed. 2014. p. 675–8.

21. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. Med Image Anal. 2017;42:60–88.

22. Sugano K. Screening of gastric cancer in Asia. Best Pract Res Clin Gastroenterol. Baillière Tindall; 2015;29:895–905.

23. Kudo T, Kakizaki S, Sohara N, et al. Analysis of ABC (D) stratification for screening patients with gastric cancer. World J Gastroenterol. Baishideng Publishing Group Inc; 2011;17:4793–8.

24. Lakhani P, Sundaram B. Deep Learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology. Radiological Society of North America; 2017;284:574–82.

25. Kim KH, Choi SH, Park S-H. Improving Arterial spin labeling by using deep learning. Radiology. Radiological Society of North America; 2018;287:658–66.

26. Larson DB, Chen MC, Lungren MP, et al. Performance of a deep-learning neural network model in assessing skeletal maturity on pediatric hand radiographs. Radiology. Radiological Society of North America; 2018;287:313–22.

27. Ishihara K, Ogawa T, Haseyama M. Classification of gastric cancer risk from X-ray images based on efficient image features related to serum Hp antibody level and serum PG levels. ITE Trans Media Technol Appl. The Institute of Image Information and Television Engineers; 2016;4:337–48.

28. Ishihara K, Ogawa T, Haseyama M. Helicobacter pylori infection detection from gastric X-ray images based on feature fusion and decision fusion. Comput Biol Med. Elsevier Ltd; 2017;84:69–78.

29. Togo R, Ishihara K, Mabe K, et al. Preliminary study of automatic gastric cancer risk classification from photofluorography. World J Gastrointest Oncol Febr World J Gastrointest Oncol. 2018;15:62–70.

Table 1. ABC (D) stratification.

	A	B	C (D)
H. pylori antibody level	−	+	+ (−)
PG levels	−	−	+

Patients with *H. pylori* antibody level ≥ 10 U/ml were classified as (+) and patients with PG I ≤ 70 ng/ml and PG I/PG II ratio ≤ 3 were classified as (+). *H. pylori* = *Helicobacter pylori*, PG = pepsinogen.

Table 2. Gastritis detection performance of our method and that of ABC (D) stratification.

	Se	Sp	Ha
Our method	0.962	0.983	0.972
ABC (D) stratification	0.925	0.998	0.960

Se = sensitivity, Sp = specificity, Ha = harmonic mean.

Figure Legends

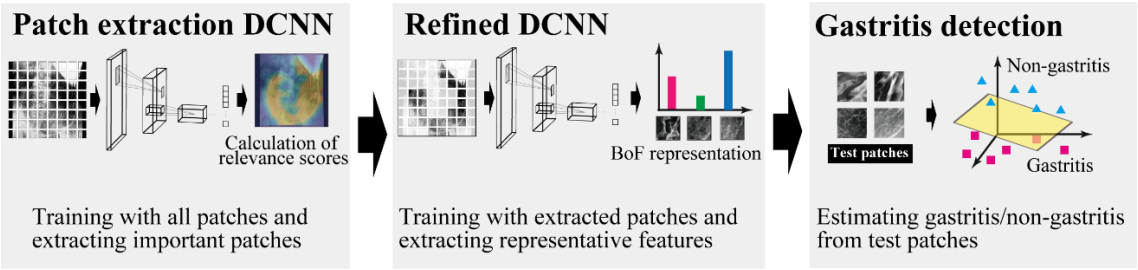


Figure 1. Target selection flowchart. UGI-XR = double contrast upper gastrointestinal barium X-ray radiography, UGI-ES = upper gastrointestinal endoscopy.

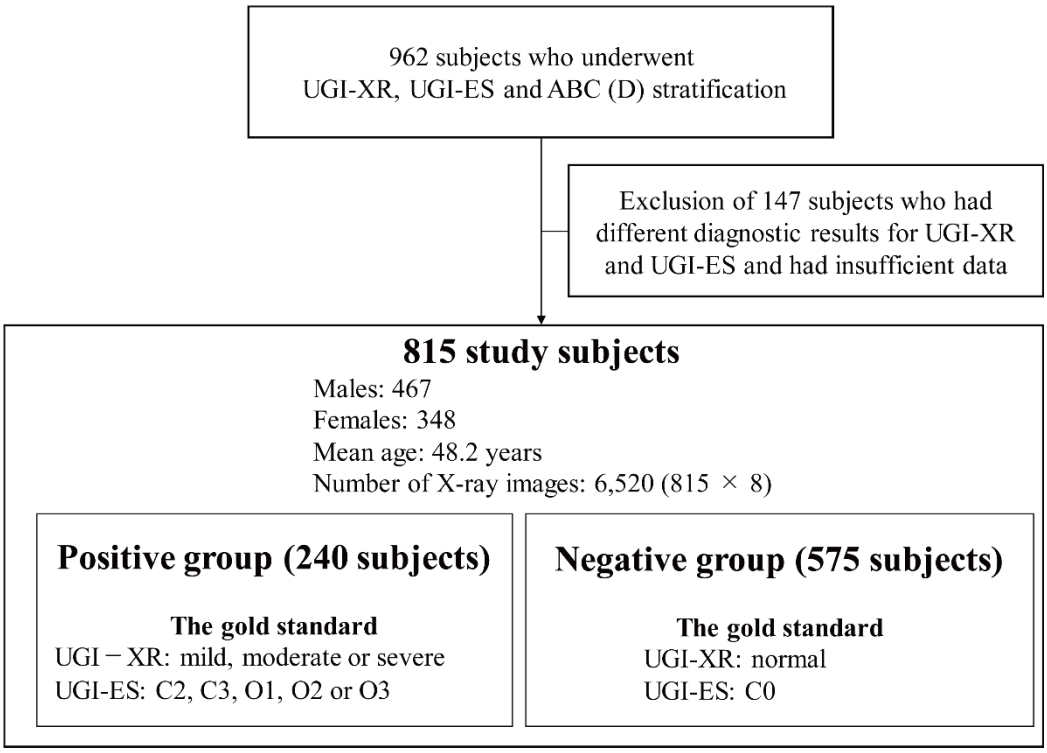


Figure 2. Overview of our gastritis detection method. DCNN = deep convolutional neural network.

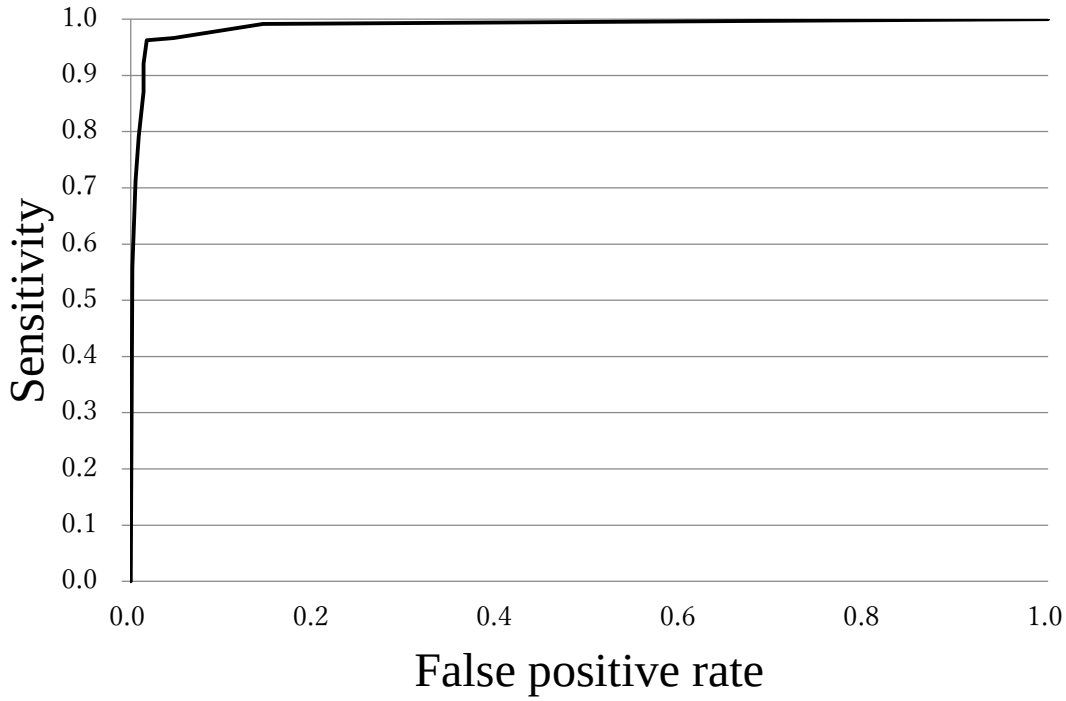
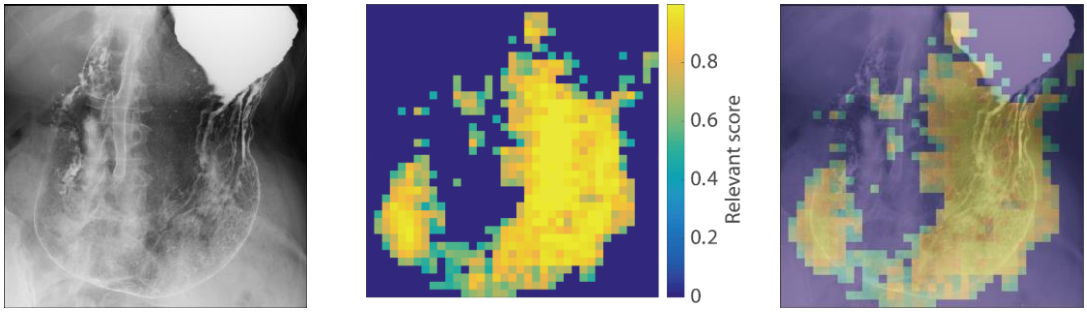
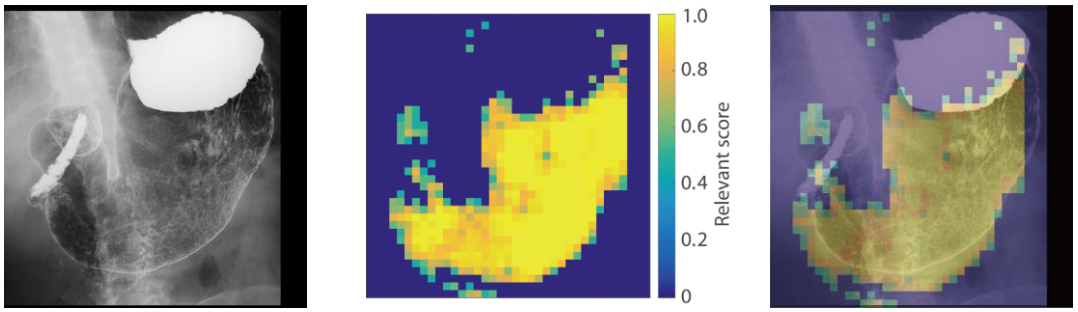


Figure 3. Receiver operating characteristic curve of our method generated by changing the threshold.



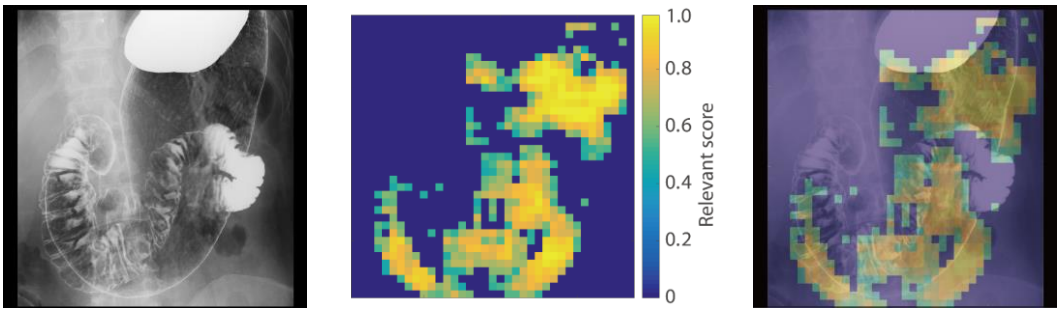
(a)-1 (a)-2 (a)-3

(a) gastritis image (ABC (D) stratification result was B).



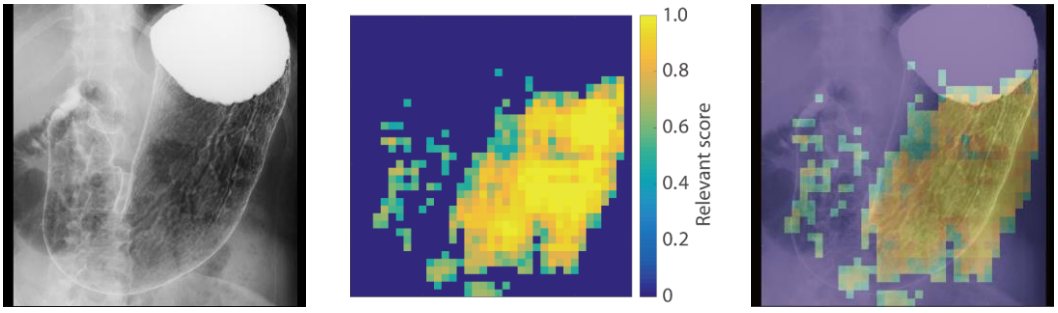
(b)-1 (b)-2 (b)-3

(b) gastritis image (ABC (D) stratification result was C).



(c)-1 (c)-2 (c)-3

(c) non-gastritis image (ABC (D) stratification result was A).



(d)-1 (d)-2 (d)-3

(d) non-gastritis image (ABC (D) stratification result was A)

Figure 4. Visualization results of relevance scores.

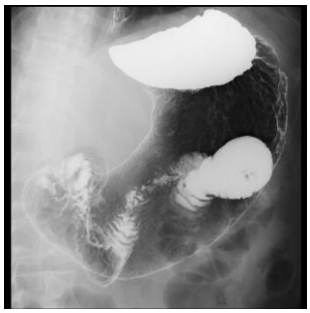
(a): (a)-1 is a sample of gastritis image (ABC (D) stratification results was B), (a)-2 is its relevance score, and (a)-3 is their composition image.

(b): (b)-1 is a sample of gastritis image (ABC (D) stratification results was C), (b)-2 is its relevance score, and (b)-3 is their composition image.

(c): (c)-1 is a sample of non-gastritis image (ABC (D) stratification results was A), (c)-2 is its relevance score, and (c)-3 is their composition image.

(d): (d)-1 is a sample of gastritis image (ABC (D) stratification results was A), (d)-2 is its relevance score, and (d)-3 is their composition image.

(a)-1

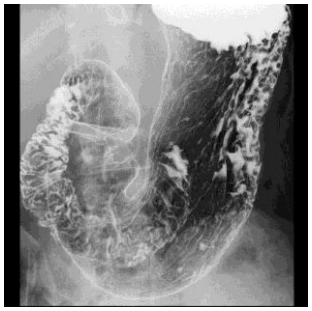


(a)-2

(a)-3

(b)-1

(b)-2



(b)-3

Figure 5. Incorrectly classified cases in our automated detection method. (a)-1, (a)-2 and (a)-3 are false negative cases. (b)-1, (b)-2 and (b)-3 are false positive cases. Blood inspection results of each case were as follows:

(a)-1: *H. pylori* antibody titer was 51.6 U/mL, PG I was 34.8 ng/ml and PG II was 18.2 ng/ml. ABC (D) stratification result was C.

(a)-2: *H. pylori* antibody titer was 63.4 U/mL, PG I was 39.0 ng/ml and PG II was 21.3 ng/ml. ABC (D) stratification result was C.

(a)-3: *H. pylori* antibody titer was 46.9 U/mL, PG I was 139.0 ng/ml and PG II was 18.8 ng/ml. ABC (D) stratification result was B.

(b)-1: *H. pylori* antibody titer was 2.3 U/mL, PG I was 47.5 ng/ml and PG II was 8.9 ng/ml. ABC (D) stratification result was A.

(b)-2: *H. pylori* antibody titer was 1.8 U/mL, PG I was 45.8 ng/ml and PG II was 5.1 ng/ml. ABC (D) stratification result was A.

(b)-3: *H. pylori* antibody titer was 0.2 U/mL, PG I was 128.0 ng/ml and PG II was 14.2 ng/ml. ABC (D) stratification result was A. *H. pylori* = *Helicobacter pylori*, PG = pepsinogen.