Received: 28 August 2023 | Accepted: 14 January 2024

DOI: 10.1002/ueg2.12551

ORIGINAL ARTICLE

ueg journal WILEY

Computer-aided diagnosis in real-time endoscopy for all stages of gastric carcinogenesis: Development and validation study

Eun Jeong Gong[1,2,3] | Chang Seok Bang[1,2,3,4] | Jae Jun Lee[3,4,5]

[1]Department of Internal Medicine, Hallym University College of Medicine, Chuncheon, Korea

[2]Institute for Liver and Digestive Diseases, Hallym University, Chuncheon, Korea

[3]Institute of New Frontier Research, Hallym University College of Medicine, Chuncheon, Korea

[4]Division of Big Data and Artificial Intelligence, Chuncheon Sacred Heart Hospital, Chuncheon, Korea

[5]Department of Anesthesiology and Pain Medicine, Hallym University College of Medicine, Chuncheon, Korea

Correspondence

Chang Seok Bang, Department of Internal Medicine, Hallym University College of Medicine, Sakju-ro 77, Chuncheon, Gangwon-do 24253, Korea.
Email: csbang@hallym.ac.kr

Funding information

Kangwon Branch of Korean Society of Gastrointestinal Endoscopy, Grant/Award Number: KSGE_KW_08; Korean Research Corporation for Helicobacter and Microbiome Grant, Grant/Award Number: KRAHM-202102003

Abstract

Objective: Using endoscopic images, we have previously developed computer-aided diagnosis models to predict the histopathology of gastric neoplasms. However, no model that categorizes every stage of gastric carcinogenesis has been published. In this study, a deep-learning-based diagnosis model was developed and validated to automatically classify all stages of gastric carcinogenesis, including atrophy and intestinal metaplasia, in endoscopy images.

Design: A total of 18,701 endoscopic images were collected retrospectively and randomly divided into train, validation, and internal-test datasets in an 8:1:1 ratio. The primary outcome was lesion-classification accuracy in six categories: normal/atrophy/intestinal metaplasia/dysplasia/early /advanced gastric cancer. External-validation of performance in the established model used 1427 novel images from other institutions that were not used in training, validation, or internal-tests.

Results: The internal-test lesion-classification accuracy was 91.2% (95% confidence interval: 89.9%–92.5%). For performance validation, the established model achieved an accuracy of 82.3% (80.3%–84.3%). The external-test per-class receiver operating characteristic in the diagnosis of atrophy and intestinal metaplasia was 93.4 ± 0% and 91.3 ± 0%, respectively.

Conclusions: The established model demonstrated high performance in the diagnosis of preneoplastic lesions (atrophy and intestinal metaplasia) as well as gastric neoplasms.

KEYWORDS

atrophy, deep learning, endoscopy, gastric cancer, gastric neoplasms, intestinal metaplasia

INTRODUCTION

Endoscopists' important skills include suspicion of lesions, early detection, and skillfully separating malignant and pre-malignant lesions from benign lesions during gastrointestinal endoscopy.[1–3] Even highly trained endoscopists cannot avoid missing neoplastic lesions during screening endoscopy in terms of lesion detection. One systematic review found that about 10% of gastric cancer screening endoscopies were missed, potentially missing one out of every 10 cancers.[4] Visual diagnosis during endoscopy for lesion classification is

also not perfect. The sensitivity and specificity for detecting gastric cancer were not perfect according to the analysis of the Korean national cancer screening program (sensitivity: 69% specificity: 96%).[5] The results of the analysis of the Japanese local gastric cancer screening program were also consistent (sensitivity: 89%, specificity: 85%).[6]

Conscientious examination with a humble attitude, examination reducing blind spots, or increasing visibility with image-enhanced endoscopy have all been advised for improving diagnostic performance in endoscopy.[7,8] However, it can be challenging to maintain a certain level of performance that is unaffected by individual skill or level of fatigue.

A single center tandem randomized study carried out in China randomly assigned 1886 upper endoscopic examinations to either routine conventional endoscopy or artificial intelligence (AI)-assisted screening endoscopy. The neoplasm miss rate during AI-assisted endoscopic examination was 6.4%. A routine screening endoscopy without AI support, however, revealed a 25.6% miss rate. The neoplasm miss rate was significantly lower in the AI-assisted group. This study demonstrated that, regardless of an endoscopist's skill level or level of fatigue, AI can be a complementary technique for their performance.[9]

We previously developed computer-aided diagnosis (CADx) models[7] for classifying the histology of gastric lesions into five classes (advanced gastric cancer [AGC], early gastric cancer [EGC], low-grade dysplasia, high-grade dysplasia, and non-neoplasm) or two classes (cancer vs. noncancer, neoplasm vs. non-neoplasm). This was followed by a model[10] that used transfer learning of pre-trained convolutional neural networks to classify the invasion-depth (mucosa-confined vs. submucosa-invaded) of gastric neoplasms from endoscopic images. We also used deep-learning to develop and validate a clinical decision support system (CDSS) for the automated detection (computer-aided detection [CADe]), diagnosis (CADx), and invasion-depth prediction (CADx) of gastric neoplasms in real-time endoscopy.[8] However, no model has been reported that classifies the entire process of gastric carcinogenesis. Furthermore, atrophy or intestinal metaplasia is a preneoplastic condition that requires close monitoring for the development of gastric cancer and testing for *Helicobacter pylori* infection (Figure 1).[11,12] Although these two conditions were categorized as normal (not a neoplasm category) in our previous studies, they should be monitored and classified differently in a clinic.[13] In this study, a deep-learning-based diagnosis model was developed and validated for the automated classification of all stages of gastric carcinogenesis, including atrophy and intestinal metaplasia, in endoscopy images.

Key summary

Summarize the established knowledge on this subject

- In this study, a deep-learning-based diagnosis model was developed and validated to automatically classify all stages of gastric carcinogenesis, including atrophy and intestinal metaplasia, in endoscopy images.

What are the significant and/or new findings of this study?

- The established computer-aided diagnosis model demonstrated high performance in the diagnosis of preneoplastic lesions such as atrophy and intestinal metaplasia as well as gastric neoplasms. This model can be implemented in the clinical practice not to miss the atrophy or intestinal metaplasia, as well as gastric neoplasms in screening upper gastrointestinal endoscopy.

MATERIALS AND METHODS

This study was approved by the institutional review board of Chuncheon Sacred Heart hospital (approval number: 2022-03-002). This study builds on previous research[7,8,10,14] by automating the classification of the entire stages of gastric carcinogenesis in endoscopic images. Because the data were collected retrospectively, the informed consent was waived. The study's schematic diagram is displayed in Figure 2. All images from the train, validation, internal-test, and external-test datasets were mutually exclusive.

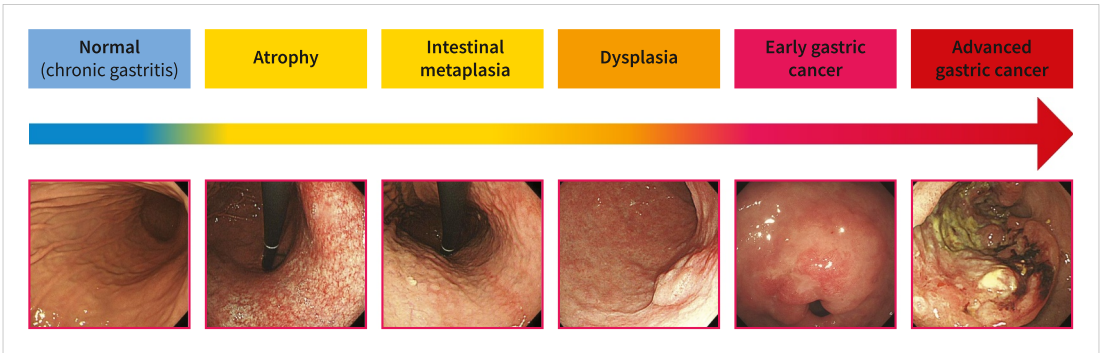

FIGURE 1 Schematic diagram of gastric carcinogenesis.

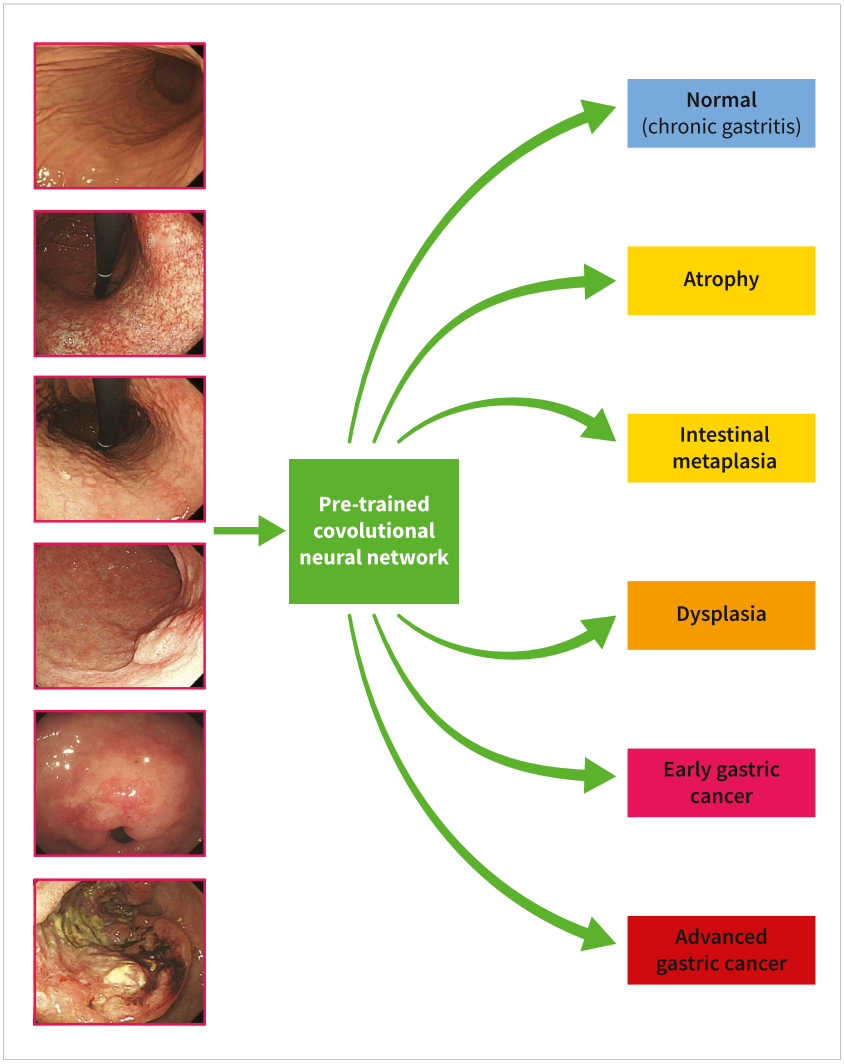

FIGURE 2 Schematic diagram for the establishment of computer-aided diagnosis model for the entire steps of gastric carcinogenesis.

Datasets

We expanded the data collection procedure used in earlier studies in order to create the CADx model.[7,8,10,14] First, between 2010 and 2022, we enrolled consecutive patients at the Hallym University Chuncheon Sacred Heart hospital with any type of gastric neoplasms identified during upper gastrointestinal endoscopy and histologically confirmed. These three categories—dysplasia, EGC, and AGC—were used to classify gastric neoplasms.

In the second step, we enrolled all patients diagnosed with atrophy or intestinal metaplasia during upper gastrointestinal endoscopy at Hallym University Chuncheon Sacred Heart hospital between 2019 and 2022. These images were assigned to one of two groups: atrophy or intestinal metaplasia. To reduce inter-observer variability and ensure accurate categorization, two expert endoscopists (C.S.B. and E.J.G.) cross-checked all enrolled images. The discussion was used to resolve the discordant categorized images.

The normal category without atrophy or intestinal metaplasia was prepared for the final step using the same procedure described above. Patients were found to be free of gastric neoplasm, atrophy, or intestinal metaplasia during upper gastrointestinal endoscopy at Hallym University Chuncheon Sacred Heart hospital between 2019 and 2022. As a result, all neoplasm categories were pathology-confirmed lesions, but atrophy, intestinal metaplasia, and normal mucosa were classified by expert endoscopists based on visual diagnosis. To reduce inter-observer variability and ensure accurate categorization, two expert endoscopists (C.S.B. and E.J.G.) cross-checked all enrolled images. Representative endoscopic images of each patient were obtained in JPEG format from the in-hospital database, with a minimum resolution of 512 × 431 pixels. Imaging software INFINITT Picture Archiving and Communication System (PACS) M6 capturing device (INFINITT Healthcare) was used. JPEG 2000 compression was used and the stored images in the PACS have a resolution of

96 or 150 dpi. Post processing was not performed on the images used in this study.

Finally, 18,701 endoscopic images were enrolled (The number of images and the number of lesions were the same) and randomly assigned to train, validation, and internal-test datasets in an 8:1:1 ratio. Randomization was done on the basis of patients rather than images.[7] Thus, lesions of the same category in a single patient were assigned in one group to either the training or test dataset. It should be noted that if a patient had lesions from multiple categories at the same time, the lesions could belong to various datasets because lesions from different categories were randomized independently.[7]

All training and internal-test examinations were performed using GIF-Q260, H260, or H290 endoscopes (Olympus Optical Co., Ltd.) in conjunction with an endoscopic video imaging system (Evis Lucera CV-260 SL or Elite CV-290; Olympus Optical Co., Ltd.). Table 1 describes the detailed distribution of the input images.

Preprocessing of training dataset

Endoscopic still-cut images contain noise information such as the date of the examination, the patients' name, age, gender, or identification number.[8] Before training, noise information was anonymized and not saved in the training dataset. To increase the amount of data, different combinations of augmentation or copies of modified images are randomly applied during training. As a result, data augmentation methods such as hue, rotation (90°), brightness, saturation, contrast, noise, horizontal or vertical flipping of included images, and image normalization with linear transformation in terms of three RGB channels were used.[7,10]

Establishment of CADx models

In this study, the no-code deep-learning tool "Neuro-X" version 3.1.1 (Neurocle Inc.) was used. Deep-learning models for image recognition and classification can be established using a software algorithm that analyzes the features of the dataset and self-discovers optimal hyperparameters, making it simple for non-experts to build the best performance models.[14] Neuro-X offers three backbone convolutional neural network architectures to choose from depending on the size of your dataset. It consists of compact, normal, and heavy architectures, with five levels of hyperparameters to adjust, as well as a choice of optimizer, decay method, batch size, epoch, and patience. We tried to obtain the best performance model possible. The entire deep-learning model development process was approached by simply clicking menus based on user-friendly graphical user interfaces in on-premise software. The training system included four RTX 3090ti graphics processing units, AMD Ryzen Threadripper PRO 5975WX 32-Core central processing units, and 512 GB RAM.

Study outcomes

The primary outcome was the established models'lesion-classification accuracy. Precision (defined as [true positive/true positive + false positive]), recall (defined as [true positive/true positive + false negative]), *F*1 score (2 precision recall/precision + recall), and per-class area under the receiver operating characteristic (AUROC) were additional performance metrics.

Performance verification for the lesion classification accuracy

To ensure the generalizability of classification performance, a performance verification (prospective validation) test was carried out using external-test datasets from other institutions. This external-test set was gathered from consecutive patients who underwent upper gastrointestinal endoscopy at Gangneung Asan Hospital between 2018 and 2020. A total of 1427 new images from 1427 patients were collected that were not used in the training or internal-testing of the established model. All external-tests were performed using GIF-Q260, H260, or H290 endoscopes (Olympus Optical Co., Ltd.) in conjunction with an endoscopic video imaging system (Evis Lucera CV-260 SL or Elite CV-290; Olympus Optical Co., Ltd.). Table 1 describes the detailed distribution of the external-test dataset.

TABLE 1 Data distribution for the establishment and test of computer-aided diagnosis model.

Number of images (number of patients)	Whole dataset	Training dataset	Validation dataset	Internal-test dataset	External-test dataset
Overall	18,701 (13,951)	14,959 (11,261)	1871 (1343)	1871 (1347)	1427 (683)
Advanced gastric cancer	1358 (542) (7.3%)	1086 (432) (7.3%)	136 (54) (7.3%)	136 (56) (7.3%)	199 (100) (13.9%)
Early gastric cancer	2208 (627) (11.8%)	1766 (511) (11.8%)	221 (60) (11.8%)	221 (56) (11.8%)	204 (59) (14.3%)
Dysplasia	2016 (837) (10.8%)	1612 (660) (10.8%)	202 (87) (10.8%)	202 (90) (10.8%)	207 (139) (14.5%)
Atrophy	4994 (3897) (26.7%)	3996 (3223) (26.7%)	499 (337) (26.7%)	499 (337) (26.7%)	286 (94) (20.0%)
Intestinal metaplasia	3787 (3710) (20.3%)	3029 (2965) (20.2%)	379 (371) (20.3%)	379 (374) (20.3%)	283 (43) (19.8%)
Normal	4338 (4338) (23.2%)	3470 (3470) (23.2%)	434 (434) (23.2%)	434 (434) (23.2%)	248 (248) (17.4%)

Attention map for explainability

In this study, a gradient-weighted class activation map (attention map) was implemented into the layer of a neural network to localize the discriminative regions in the given images to determine the specific class. An attention map was created and analyzed for each external-test image.[15]

RESULTS

Characteristics of the dataset

A total of 18,701 endoscopic images were collected retrospectively and randomly divided into train, validation, and internal-test datasets in an 8:1:1 ratio. The proportion of "atrophy" in the total images was highest (26.7% [4994/18,701]), followed by "normal" (23.2% [4338/18,701]) and "intestinal metaplasia" (20.3% [3787/18,701]). In terms of neoplastic lesions, "EGC" had the highest proportion (11.8% [2208/18,701]), followed by "dysplasia" (10.8% [2016/18,701]) and "AGC" (7.3% [1358/18,701]).

A total of 1427 novel images with proportions that reflected the unique characteristics of the own institution were collected for the external-test dataset. The proportion of "atrophy" was the highest (20% [286/1427]), followed by "intestinal metaplasia" (19.8% [283/1427]) and "normal" (17.4% [248/1427]). The proportion of "dysplasia" in the neoplasm categories was highest (14.5% [207/1427]), followed by "EGC" (14.3% [204/1427]) and "AGC" (13.9% [199/1427]). Table 1 describes the number and distribution of each category in the dataset.

Training parameters in the establishment of CADx model

The on-premise software's own neural network structure was used with Adam optimizer, Cosine learning rate decay method (decay ends at 23,400 steps, initial learning rate 0.002), batch size of 40, epoch of 100, and patience of 30 to establish the CADx model. Inference time was 6.18 ms and loss function was categorical crossenteropy. The total time spent training was 5 h and 19 min. The validation loss was 0.0708 and the train loss was 0.3476 (at epoch 40).

Internal-test performance

The established model had an accuracy of 91.2% (95% confidence interval, 89.9%–92.5%), the precision of 88.4% (86.9%–89.9%), recall

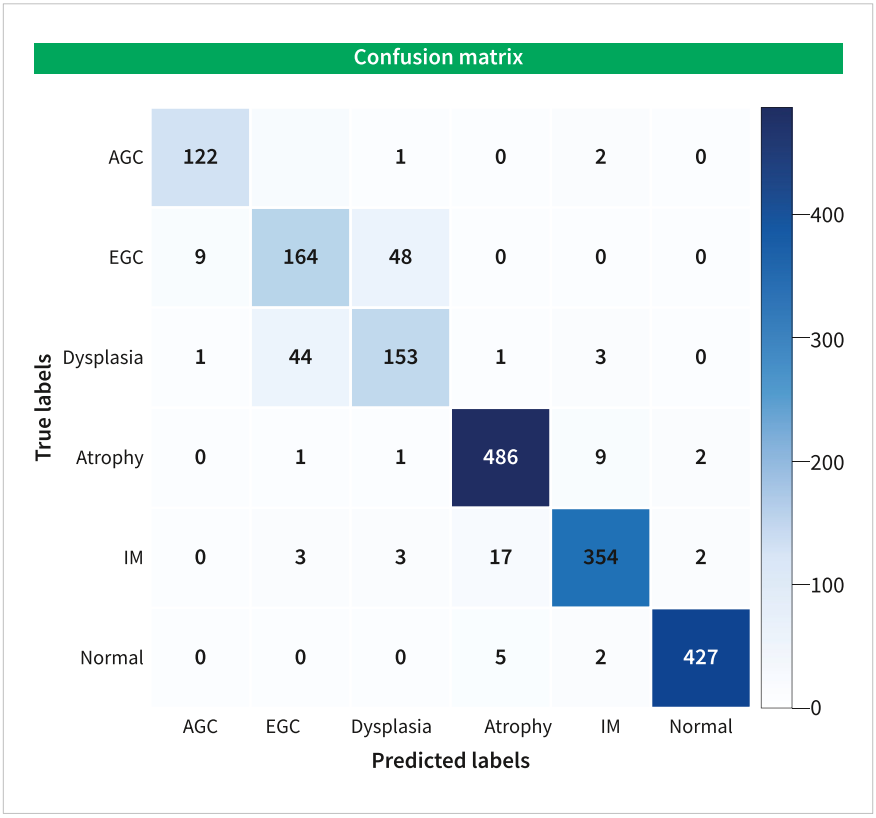

FIGURE 3 Confusion matrix for the computer-aided diagnosis model in the internal-test.

TABLE 2 A summary of the performance of computer-aided diagnosis model in white-light endoscopy images.

% (95% confidence interval)	Accuracy	Precision	Recall	*F*1 score
Internal-test performance (n = 1871)	91.2% (89.9%–92.5%)	88.4% (86.9%–89.9%)	88.1% (86.6%–89.6%)	88.2% (86.7%–89.7%)
External-test performance (n = 1427)	82.3% (80.3%–84.3%)	83% (81.1%–84.9%)	82.3% (80.3%–84.3%)	82.6% (80.6%–84.6%)

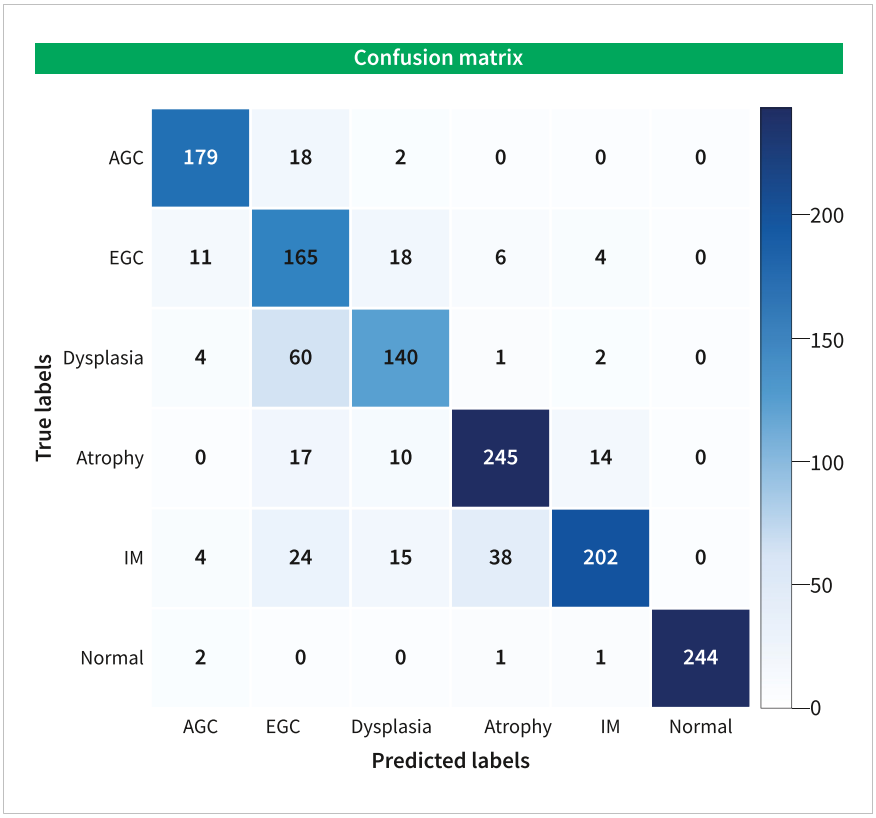

FIGURE 4 Confusion matrix for the computer-aided diagnosis model in the external-test.

of 88.1% (86.6–89.6), and *F*1 score of 88.2% (86.7%–89.7%) in internal-tests. The per-class AUROC is 97.3% for AGC, 94.1% for EGC, 93.4% for dysplasia, 97.3% for atrophy, 96.9% for intestinal metaplasia, and 99.5% for the normal category. Figure 3 demonstrates the confusion matrix for internal-test performance, and Table 2 describes the detailed performance of the established CADx model.

External-test performance

The established model had an accuracy of 82.3% (95% confidence interval, 80.3%–84.3%), the precision of 83% (81.1%–84.9%), recall of 82.3% (80.3–84.3), and *F*1 score of 82.6% (80.6%–84.6%) in external-tests. The per-class AUROC is 96.7% for AGC, 91.8% for EGC, 93.4% for dysplasia, 93.4% for atrophy, 91.3% for intestinal metaplasia, and 99.3% for the normal category. Figure 4 demonstrates the confusion matrix for internal-test performance, and Table 2 describes the detailed performance of the established CADx model.

Attention map

Figure S1 shows representative cases of neoplastic lesions, while Figure S2 shows representative cases of atrophy or intestinal metaplastic lesions that were correctly determined region of interest by an established CADx model. The white dotted line represents the ground truth region of interest, which was well matched with the CADx model's determination.

Figure S3 demonstrates the incorrectly determined cases of the CADx model. Although the CADx model incorrectly identified a region of interest (by focusing on only a portion of the lesion), the characteristic area of the lesion was well noted by an established CADx model in most cases.

DISCUSSION

We developed and validated a CADx model for automated diagnosis of the entire steps of gastric carcinogenesis in upper gastrointestinal endoscopy. Gastric atrophy and intestinal metaplasia, according to Correa's hypothesis, are preneoplastic conditions that should be monitored for the development of gastric neoplasms.[16,17] In patients with atrophy or intestinal metaplasia, the American Gastroenterology Association recommended *H. pylori* testing and eradication to prevent the progression of gastric carcinogenesis.[18,19] They must also be monitored on a regular basis. Even after *H. pylori* eradication, patients with severe atrophy or intestinal metaplasia (OLGA, OLGIM stage III-IV) are at high risk of developing gastric neoplasms.[20]

CDSS for the automated detection and diagnosis of gastric neoplasms in real endoscopic procedures has already been established and validated by our group. However, because atrophy or intestinal metaplasia were not classified as neoplasms, they were classified as "normal" at the time of CDSS development. To compensate for the need for individualized risk stratification, we added two new classes to the total dataset and extended data preparation steps to ensure that the model works properly in real-world scenarios. Because the diagnostic capability of our previous CADx models for neoplastic lesions was already qualified, we concentrated on the diagnostic capability of preneoplastic conditions such as atrophy and intestinal metaplasia. As a result, the proportion of "atrophy" (26.7%) or "intestinal metaplasia" (20.3%) was twice that of neoplastic lesions. We also conducted a pilot study to determine the best distribution of input training data with the best performance, and the model with an even distribution of each category demonstrated lower diagnostic accuracy than the currently established model (approximately half of the "atrophy" or "intestinal metaplasia" images were used).

Priorities for AI in gastrointestinal endoscopy were recommended by the American Society of Gastrointestinal Endoscopy.[21] CADe of gastric cancer precursor lesions, including atrophy and intestinal metaplasia, was one of the recommended priorities.[21] Although the current work is not a CADe but rather a CADx, we plan to expand our research by developing a CADe model and combining CADe and CADx into a single CDSS. According to the European Society of Gastrointestinal Endoscopy, CADx of gastric preneoplastic conditions should be one of the top priorities for AI use.[22] Real-time AI-assisted diagnosis of gastric preneoplastic conditions with diagnostic accuracy measurement was recommended, and our current study used the same task.

Previous research on this topic has shown that CADe models have comparable or even better diagnostic performance than endoscopists and have clinical utility. Guimares et al. developed an atrophy detection model and demonstrated 93% accuracy in an independent data set, outperforming expert endoscopists.[23] Zhang et al. also established CADe of gastric atrophy and demonstrated 94% accuracy, sensitivity, and specificity, which was higher than that of endoscopic experts.[24] Zhao et al. also created a CADe of atrophy and carried out a prospective cohort study. They demonstrated that using the CADe model improved the diagnosis rate of gastric atrophy when compared to not using the CADe model.[25] Luo et al. developed a CADe model for gastric atrophy that demonstrated comparable detection performance to endoscopists.[26] However, in real-world clinical practice, this type of dataset is meaningless. These studies created a training dataset with only atrophy and no atrophy categories. In these studies, the categories of gastric cancer, normal, and intestinal metaplasia are not taken into account. Because only atrophy versus no-atrophy discrimination is possible, these models cannot be used in clinical settings.

Siripoppohn et al. established a real-time semantic segmentation model in terms of intestinal metaplasia.[27] Although the diagnostic performance in this study was excellent, this model was not a CADe or CADx model; rather, only gastric intestinal metaplasia could be segmented in real time.

This study developed and validated a CADx model for the automated diagnosis of all stages of gastric carcinogenesis during upper gastrointestinal endoscopy. The diagnostic performance demonstrated clinical utility (internal-test lesion-classification accuracy of 91.2%, external-test accuracy of 82.3%, external-test per-class AUROC of atrophy and intestinal metaplasia of 93.4 and 91.3, respectively). The number of correctly identified regions of interest for external-test images was comparable between expert endoscopist and established model. To the best of our knowledge, this is the first study to establish and validate the diagnostic performance of a CADx model for gastric atrophy and intestinal metaplasia, taking into account all stages of gastric carcinogenesis.

Despite the promising results mentioned above, several unavoidable limitations were discovered. First, the training dataset was obtained from a single institution, which may indicate a selection or spectrum bias. Medical AI models developed from a single institution typically have limitations for widespread implementation due to the unique characteristics of patients in each institution, highlighting the importance of external-testing.[28] To compensate for this flaw, we conducted stringent validations and included images from other institutions. Second, a CADx model without a CADe model was created. In general, the flow of conventional screening endoscopy is lesion detection and classification. However, we only created a CADx model before creating a CADe model. We created a diagnosis model first because we determined the distribution of training data while creating the diagnosis model, and in the follow-up study, we plan to create a detection model based on this training data and integrate it into one CDSS. Third, despite the fact that we used gradient-weighted class activation maps for our analysis, these maps have their own drawbacks, including the inability to localize multiple instances of an object in an image due to partial derivatives premise and inaccurate localization of the heatmap with respect to the coverage of the class area.[29] Fourth, we used a JPEG baseline format for model establishment. This compression standard is typically "lossy" and includes several user-defined settings that affect image quality. Although JPEG was the only format that could be collected in our study, it may have caused a bias in image quality. Further

research using only TIFF or PNG file formats would avoid this type of bias.[7] Fourth, several lesions from the same patient may be introduced into different sets. Although the study focuses on lesions, there is still a bias if multiple lesions from the same patient are included in both training and test sets. However, including only one image from one lesion could lead to insufficient model performance. Fifth, the normal category in our study was classified by expert endoscopists based on visual diagnosis, which may introduce potential misclassification. Due to the retrospective nature of this study, histological confirmation of the normal category was not feasible. This limitation is important to consider when interpreting our findings. Sixth, because images found to be free of gastric neoplasm, atrophy, or intestinal metaplasia were collected more recently than the average positive sample, this could result in a difference in image quality between the negative and positive samples. Seventh, the current study has the potential for information leakage. Information leakage refers to a situation where information outside the training data is used to establish the model. This can lead to inaccurate performance of the mode being built. We tried to limit this risk by using two independent test sets that are not used for hyperparameter tuning.

In conclusion, the established computer-aided diagnosis model demonstrated high performance in the diagnosis of preneoplastic lesions such as atrophy and intestinal metaplasia as well as gastric neoplasms.

AUTHOR CONTRIBUTIONS

Conceptualization: Chang Seok Bang. *Data curation*: Eun Jeong Gong, Chang Seok Bang, Jae Jun Lee. *Formal analysis*: Eun Jeong Gong, Chang Seok Bang, Jae Jun Lee. *Funding acquisition*: Chang Seok Bang. *Investigation*: Eun Jeong Gong, Chang Seok Bang, Jae Jun Lee. *Methodology*: Chang Seok Bang. *Project administration*: Chang Seok Bang. *Resources*, Eun Jeong Gong, Chang Seok Bang, Jae Jun Lee. *Supervision*: Chang Seok Bang. *Writing - original draft*: Eun Jeong Gong. *Writing - review and editing*: Chang Seok Bang.

ACKNOWLEDGMENTS

This work was supported by the Korean Research Corporation for Helicobacter and Microbiome Grant (Grant Number, KRAHM-202102003) and Kangwon Branch of Korean Society of Gastrointestinal Endoscopy (KSGE_KW_08).

CONFLICT OF INTEREST STATEMENT

The authors disclose no conflicts of interest. The remaining authors declare that there are no conflicts of interest.

DATA AVAILABILITY STATEMENT

All the data are accessible and available upon request by the corresponding author.

DECLARATIONS

Not applicable.

ACCESS TO DATA

All investigators have access to the final dataset.

IRB APPROVAL NUMBER

2022-03-002.

ORCID

Eun Jeong Gong https://orcid.org/0000-0003-3996-3472
Chang Seok Bang https://orcid.org/0000-0003-4908-5431

REFERENCES

1. Bang CS. Deep learning in upper gastrointestinal disorders: status and future perspectives. Korean J Gastroenterol. 2020;75(3):120–31. https://doi.org/10.4166/kjg.2020.75.3.120
2. Bang CS. Artificial intelligence in the analysis of upper gastrointestinal disorders. Korean J Helicobacter Upper Gastrointest Res. 2021;21(4):300–10. https://doi.org/10.7704/kjhugr.2021.0030
3. Cho BJ, Bang CS. Artificial intelligence for the determination of a management strategy for diminutive colorectal polyps: hype, hope, or help. Am J Gastroenterol. 2020;115(1):70–2. https://doi.org/10.14309/ajg.0000000000000476
4. Pimenta-Melo AR, Monteiro-Soares M, Libânio D, Dinis-Ribeiro M. Missing rate for gastric cancer during upper gastrointestinal endoscopy: a systematic review and meta-analysis. Eur J Gastroenterol Hepatol. 2016;28:1041–9. https://doi.org/10.1097/meg.0000000000000657
5. Choi KS, Jun JK, Park EC, Park S, Jung KW, Han MA, et al. Performance of different gastric cancer screening methods in Korea: a population-based study. PLoS One. 2012;7(11):e50041. https://doi.org/10.1371/journal.pone.0050041
6. Hamashima C, Okamoto M, Shabana M, Osaki Y, Kishimoto T. Sensitivity of endoscopic screening for gastric cancer by the incidence method. Int J Cancer. 2013;133(3):653–9. https://doi.org/10.1002/ijc.28065
7. Cho BJ, Bang CS, Park SW, Yang YJ, Seo SI, Lim H, et al. Automated classification of gastric neoplasms in endoscopic images using a convolutional neural network. Endoscopy. 2019;51(12):1121–9. https://doi.org/10.1055/a-0981-6133
8. Gong EJ, Bang CS, Lee JJ, Baik GH, Lim H, Jeong JH, et al. Deep learning-based clinical decision support system for gastric neoplasms in real-time endoscopy: development and validation study. Endoscopy. 2023;55(08):701–8. https://doi.org/10.1055/a-2031-0691
9. Wu L, Shang R, Sharma P, Zhou W, Liu J, Yao L, et al. Effect of a deep learning-based system on the miss rate of gastric neoplasms during upper gastrointestinal endoscopy: a single-centre, tandem, randomised controlled trial. Lancet Gastroenterol Hepatol. 2021;6(9):700–8. https://doi.org/10.1016/s2468-1253(21)00216-8
10. Cho BJ, Bang CS, Lee JJ, Seo CW, Kim JH. Prediction of submucosal invasion for gastric neoplasms in endoscopic images using deep-learning. J Clin Med. 2020;9(6):1858. https://doi.org/10.3390/jcm9061858
11. Bang CS, Lee JJ, Baik GH. Artificial intelligence for the prediction of *Helicobacter pylori* infection in endoscopic images: systematic review and meta-analysis of diagnostic test accuracy. J Med Internet Res. 2020;22(9):e21983. https://doi.org/10.2196/21983
12. Bang CS, Lim H, Jeong HM, Shin WG, Choi JH, Soh JS, et al. Amoxicillin or tetracycline in bismuth-containing quadruple therapy as first-line treatment for *Helicobacter pylori* infection. Gut Microb. 2020;11(5):1314–23. https://doi.org/10.1080/19490976.2020.1754118

13. Bang CS, Baik GH. Attempts to enhance the eradication rate of *Helicobacter pylori* infection. World J Gastroenterol. 2014;20(18):5252–62. https://doi.org/10.3748/wjg.v20.i18.5252
14. Bang CS, Lim H, Jeong HM, Hwang SH. Use of endoscopic images in the prediction of submucosal invasion of gastric neoplasms: automated deep learning model development and usability study. J Med Internet Res. 2021;23(4):e25167. https://doi.org/10.2196/25167
15. Gong EJ, Bang CS, Jung K, Kim SJ, Kim JW, Seo SI, et al. Deep-learning for the diagnosis of esophageal cancers and precursor lesions in endoscopic images: a model establishment and nationwide multicenter performance verification study. J Pers Med. 2022;12(7):1052. https://doi.org/10.3390/jpm12071052
16. Correa P, Haenszel W, Cuello C, Tannenbaum S, Archer M. A model for gastric cancer epidemiology. Lancet. 1975;2:58–60.
17. Correa P, Piazuelo MB. The gastric precancerous cascade. J Dig Dis. 2012;13(1):2–9. https://doi.org/10.1111/j.1751-2980.2011.00550.x
18. Shah SC, Piazuelo MB, Kuipers EJ, Li D. AGA clinical practice update on the diagnosis and management of atrophic gastritis: expert review. Gastroenterology. 2021;161(4):1325–32.e7. https://doi.org/10.1053/j.gastro.2021.06.078
19. Gupta S, Li D, El Serag HB, Davitkov P, Altayar O, Sultan S, et al. AGA clinical practice guidelines on management of gastric intestinal metaplasia. Gastroenterology 2020;158(3):693–702. https://doi.org/10.1053/j.gastro.2019.12.003
20. Malfertheiner P, Megraud F, Rokkas T, Gisbert JP, Liou JM, Schulz C, et al. Management of *Helicobacter pylori* infection: the Maastricht VI/Florence consensus report. Gut. 2022;71(9):1724–62. https://doi.org/10.1136/gutjnl-2022-327745
21. Berzin TM, Parasa S, Wallace MB, Gross SA, Repici A, Sharma P. Position statement on priorities for artificial intelligence in GI endoscopy: a report by the ASGE Task Force. Gastrointest Endosc. 2020;92(4):951–9. https://doi.org/10.1016/j.gie.2020.06.035
22. Messmann H, Bisschops R, Antonelli G, Libânio D, Sinonquel P, Abdelrahim M, et al. Expected value of artificial intelligence in gastrointestinal endoscopy: European Society of Gastrointestinal Endoscopy (ESGE) Position Statement. Endoscopy. 2022;54(12):1211–31. https://doi.org/10.1055/a-1950-5694
23. Guimarães P, Keller A, Fehlmann T, Lammert F, Casper M. Deep-learning based detection of gastric precancerous conditions. Gut. 2020;69(1):4–6. https://doi.org/10.1136/gutjnl-2019-319347
24. Zhang Y, Li F, Yuan F, Zhang K, Huo L, Dong Z, et al. Diagnosing chronic atrophic gastritis by gastroscopy using artificial intelligence. Dig Liver Dis. 2020;52(5):566–72. https://doi.org/10.1016/j.dld.2019.12.146
25. Zhao Q, Chi T. Deep learning model can improve the diagnosis rate of endoscopic chronic atrophic gastritis: a prospective cohort study. BMC Gastroenterol. 2022;22(1):133. https://doi.org/10.1186/s12876-022-02212-1
26. Luo J, Cao S, Ding N, Liao X, Peng L, Xu C. A deep learning method to assist with chronic atrophic gastritis diagnosis using white light images. Dig Liver Dis. 2022;54(11):1513–9. https://doi.org/10.1016/j.dld.2022.04.025
27. Siripoppohn V, Pittayanon R, Tiankanon K, Faknak N, Sanpavat A, Klaikaew N, et al. Real-time semantic segmentation of gastric intestinal metaplasia using a deep learning approach. Clin Endosc. 2022;55(3):390–400. https://doi.org/10.5946/ce.2022.005
28. Yang CB, Kim SH, Lim YJ. Preparation of image databases for artificial intelligence algorithm development in gastrointestinal endoscopy. Clin Endosc. 2022;55(5):594–604. https://doi.org/10.5946/ce.2021.229
29. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision. Venice: IEEE; 2017. p. 618–26.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Gong EJ, Bang CS, Lee JJ. Computer-aided diagnosis in real-time endoscopy for all stages of gastric carcinogenesis: development and validation study. United European Gastroenterol J. 2024;12(4):487–95. https://doi.org/10.1002/ueg2.12551