

Diagnostic Performance of Endoscopic Ultrasound-Artificial Intelligence Using Deep Learning Analysis of Gallbladder Polypoid Lesions: A Feasibility Study

Sung Ill Jang

Department of Internal Medicine, Gangnam Severance Hospital, Yonsei University College of Medicine

Young Jae Kim

Department of Biomedical Engineering, Gachon University College of Health Science

Eui Joo Kim

Department of Internal Medicine, Gil Medical Center, Gachon University College of Medicine

Huapyong Kang

Department of Internal Medicine, Gil Medical Center, Gachon University College of Medicine

Seung Jin Shon

Department of Internal Medicine, Gangnam Severance Hospital, Yonsei University College of Medicine

Yu Jin Seol

Department of Biomedical Engineering, Gachon University College of Health Science

Dong Ki Lee

Department of Internal Medicine, Gangnam Severance Hospital, Yonsei University College of Medicine

Kwang Gi Kim

Department of Biomedical Engineering, Gachon University College of Health Science

Jae Hee Cho (✉ JHCHO9328@yuhs.ac)

Department of Internal Medicine, Gangnam Severance Hospital, Yonsei University College of Medicine

Research Article

Keywords: Artificial intelligence, Deep learning, Endosonography, Gallbladder disease, Polyps

Posted Date: February 1st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-149275/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Diagnostic performance of endoscopic ultrasound-artificial intelligence using deep learning analysis of gallbladder polypoid lesions: a feasibility study

Sung Ill Jang¹ MD, PhD*, Young Jae Kim² PhD*, Eui Joo Kim³ MD, Huapyong Kang³ MD,
Seung Jin Shon¹ MD, Yu Jin Seol² SD, Dong Ki Lee¹ MD, PhD, Kwang Gi Kim² PhD, Jae
Hee Cho¹ MD, PhD

¹Department of Internal Medicine, Gangnam Severance Hospital, Yonsei University College of Medicine, Seoul, Korea

²Department of Biomedical Engineering, Gachon University College of Health Science, Incheon, Korea

³Department of Internal Medicine, Gil Medical Center, Gachon University College of Medicine, Incheon, Republic of Korea

* These authors contributed equally as first authors.

Correspondence: Jae Hee Cho, MD, PhD

Department of Internal Medicine, Gangnam Severance Hospital, Yonsei University College of Medicine, 211 Eonjuro, Gangnam-gu, 06273, Seoul, Korea

Phone: 82-2-2019-3310

Fax: 82-2-3463-3882

E-mail: JHCHO9328@yuhs.ac

Co-correspondence: Kwang Gi Kim, PhD

Department of Biomedical Engineering, Gachon University College of Health Science, 38-13

Dokjeom-ro 3beon-gil, Namdong-gu, 21565, Incheon, Korea

Phone: 82-32-458-2770

Fax: 82-32-460-2361

E-mail: kimkg@gachon.ac.kr

Abstracts

Endoscopic ultrasound (EUS) is the most accurate diagnostic modality for polypoid lesions of the gallbladder (GB), but is limited by subjective interpretation. We evaluated the diagnostic performance of deep learning-based artificial intelligence (AI) in differentiating polypoid lesions using EUS images. The diagnostic performance of the EUS-AI system with ResNet50 architecture was evaluated via three processes: training, internal validation, and testing. The diagnostic performance was also verified using an external validation cohort and compared with the performance of EUS endoscopists. In the AI development cohort, the diagnostic performance of EUS-AI including sensitivity, specificity, positive predictive value, negative predictive value and accuracy. For the differential diagnosis of neoplastic and non-neoplastic GB polyps, these values for EUS-AI were 77.8%, 91.6%, 57.9%, 96.5%, and 89.8%, respectively. In the external validation cohort, the differential diagnosis of neoplastic and non-neoplastic GB polyps, these values were 60.3%, 77.4%, 36.2%, 90.2%, and 74.4%, respectively, for EUS-AI; they were 74.2%, 44.9%, 75.4%, 46.2%, and 65.3%, respectively, for the endoscopists. The accuracy of the EUS-AI was between the accuracies of mid-level (66.7%) and expert EUS endoscopists (77.5%). This EUS-AI system showed favorable performance for the diagnosis of neoplastic GB polyps, with a performance comparable to that of EUS endoscopists.

Keywords: Artificial intelligence, Deep learning, Endosonography, Gallbladder disease, Polyps

Introduction

Polypoid lesions of the gallbladder (GB) are abnormal elevated lesions of the GB mucosal wall into the GB lumen and encompass a wide variety of pathologies. They can be classified into non-neoplastic (cholesterol polyps, inflammatory polyps, and adenomyomatosis) or neoplastic types (adenoma and adenocarcinoma),¹ as well as benign or malignant (adenocarcinoma). Although the widespread use of ultrasonography (US) has improved detection of a variety of GB diseases, differential diagnosis of GB polyps, particularly small lesions (≤ 20 mm in diameter), is difficult because GB lesions cannot be biopsied.²

Currently, endoscopic ultrasound (EUS) is considered superior to conventional US in terms of GB examination. EUS uses a high-frequency transducer to observe GB polyps at high resolution, which can improve the discrimination of neoplastic GB polyps and aid in staging of GB cancers.³ In addition, it is useful to differentiate neoplastic and non-neoplastic GB polyps based on the presence of typical EUS features, such as hypoechoic foci, hyperechoic spot, microcyst, and Doppler flow, as well as complex scoring systems.^{2,4-6} Contrast enhanced harmonic EUS can discriminate between adenomatous and cholesterol GB polyps based on enhancement patterns,⁷ and increase the sensitivity in identification of malignant GB polyps.⁸ However, the interpretation of EUS and contrast enhanced harmonic-EUS images by visual assessments is subjective; therefore, an objective analysis technique for discriminating GB diseases is required. Artificial intelligence (AI) is a mathematical prediction technique involving automated learning and recognition data patterns. Deep learning is an AI algorithm and an advanced type of machine learning method that uses neural networks⁹. Remarkable advancements have recently been made in increasing the efficacy of AI and decreasing the cost associated with diagnosis of various gastrointestinal lesions.¹⁰⁻¹² However, no studies have used

AI to differentiate among various GB polyps. Thus, we evaluated the diagnostic performance of AI for differentiating GB polyps using EUS images.

Results

Characteristics of patients and polypoid lesions of gallbladder in AI development cohort

Table 1 lists the baseline characteristics of patients with GB lesions, including GB polyps and gallstones, in the AI development cohort. The mean patient age was 53.3 ± 13.8 years, and 344 (56.3%) patients were women. When limited to GB polyps in this cohort, cholesterol polyps were most common (322/467 patients, 68.9%); neoplastic GB polyps were found in 99/467 patients (21.2%). Of the 1,039 EUS images, 145 were neoplastic GB polyps, 691 were non-neoplastic GB polyps, and 203 were gallstones. Histopathological diagnosis revealed that the mean maximum diameters were 15.1 ± 4.2 mm for neoplastic GB polyps, 10.6 ± 3.1 mm for non-neoplastic GB polyps, and 11.1 ± 4.0 mm for gallstones. The neoplastic GB polyps were significantly larger than non-neoplastic GB polyps and gallstones ($p < 0.001$, $p < 0.001$), but there was no difference in size between non-neoplastic GB polyps and gallstones ($p = 0.20$). However, current guidelines suggest cholecystectomy for GB polyps > 10 mm¹³; therefore, all patients included in this cohort underwent surgery, despite significant differences in size.

Diagnostic performance of EUS-AI system using the AI development cohort.

The diagnostic capability of the EUS-AI system was verified using the test data set in the AI development cohort (Table 2). In diagnostic step 1, for differentiation of GB polyps and gallstones, sensitivity was 96.3%, specificity was 92.9%, PPV was 98.4%, NPV was 79.5%, and accuracy was 95.7%. The AUROC of step 1 was 0.99 (95% confidence interval [CI], 0.98–0.99). In diagnostic step 2, for differentiation of neoplastic and non-neoplastic GB polyps,

sensitivity was 77.8%, specificity was 91.6%, PPV was 57.9%, NPV was 96.5%, and accuracy was 89.8%. The AUROC of step 2 was 0.89 (95% CI, 0.87–0.91). In diagnostic step 3, for differentiation of adenocarcinoma and adenomatous GB polyps, sensitivity was 85.6%, specificity was 75.0%, PPV was 87.4%, NPV was 72.0%, and accuracy was 82.1%. The AUROC of step 3 was 0.91 (95% CI, 0.82–0.99).

Comparison of diagnostic performance between humans and EUS-AI system using external validation cohort

Because we observed encouraging diagnostic performance results with the EUS-AI system in the AI development cohort, we compared human and AI diagnostic abilities in the external validation cohort. Diagnostic performances, including sensitivity, specificity, PPV, NPV, and accuracy (Table 3), were as follows. In diagnostic step 1, these values were 63.6%, 95.8%, 70.7%, 94.5%, and 91.6%, respectively, for EUS-AI; they were 74.7%, 97.9%, 86.1%, 96.2%, and 94.9%, respectively, for the endoscopists. In diagnostic step 2, these values were 60.3%, 77.4%, 36.2%, 90.2%, and 74.4%, respectively, for EUS-AI; they were 74.2%, 44.9%, 75.4%, 46.2%, and 65.3%, respectively, for the endoscopists. For diagnostic step 3, these values were 65.3%, 60.6%, 78.3%, 44.4%, and 63.8%, respectively, for EUS-AI; they were 51.8%, 21.8%, 32.6%, 40.3%, and 34.7%, respectively, for the endoscopists. The AUROCs of diagnostic probability for GB polyps, neoplastic GB polyps, and adenocarcinomas via the EUS-AI system were 0.95 (95% CI, 0.93–0.96), 0.69 (95% CI, 0.68–0.70), and 0.63 (95% CI, 0.57–0.7), respectively (Fig. 2).

Considering proficiency and experience with EUS, the accuracy of the EUS-AI system was between the accuracies of mid-level (66.7%) and expert EUS endoscopists (77.5%) for the

differential diagnosis of neoplastic polyps and non-neoplastic polyps (Fig. 3). The accuracy of diagnosing adenocarcinoma via EUS-AI was 63.5%, which was higher than all levels of EUS endoscopists (16.7–49.2%). Furthermore, interobserver agreement between the endoscopists and EUS-AI was fair ($k = 0.22$). Among the endoscopists, interobserver agreement was relatively high for the expert EUS endoscopists ($k = 0.28$), but poor for the mid-level ($k = 0.19$) and junior endoscopists ($k = 0.01$).

Discussion

In this pilot study, we found that a deep learning-based EUS-AI algorithm was valuable for the differential diagnosis of GB polyps. The intuitive diagnosis of GB polyps involves the following steps: identification of GB polyps and exclusion of gallstones (step 1), differentiation of neoplastic and non-neoplastic polyps among GB polyps (step 2), and identification of adenocarcinoma among neoplastic polyps (step 3). These diagnostic steps are based on differences in treatment, depending on the diagnosis. Notably, laparoscopic cholecystectomy is commonly performed for GB polyps > 10 mm in size; approximately one-third of those polyps are adenomatous.¹⁴ Furthermore, it is important to detect adenocarcinomas among neoplastic GB polyps; GB cancer, which invades the muscular layer, may require extended cholecystectomy. In our study, the sensitivity, specificity, PPV, NPV, and accuracy were calculated for these diagnostic steps of the EUS-AI system. In the AI development cohort, the respective AUROCs of steps 1, 2, and 3 were 0.99 (95% CI 0.98–0.99), 0.89 (95% CI 0.87–0.91), and 0.91 (95% CI 0.82–0.99); these were very encouraging data, indicating that the EUS-AI system is useful for objective diagnosis.

Because further verification of the usefulness of the EUS-AI system is required, we performed external validation to compare diagnostic performances using a multicenter validation cohort of 83 patients with GB lesions. In this external validation cohort, the AUROCs of the EUS-AI system, in terms of the diagnostic probabilities of step 1 (identification of GB polyps and exclusion of gallstones), step 2 (differentiation of neoplastic and non-neoplastic GB polyps), and step 3 (diagnosis of adenocarcinoma among neoplastic GB polyps), were 0.95 (95% CI, 0.93–0.96), 0.69 (95% CI, 0.68–0.70), and 0.63 (95% CI, 0.57–0.7), respectively. This reduction in diagnostic capacity of steps 2 and 3, compared with that of the AI development cohort, presumably occurred for two reasons: 1) the number of EUS images in the AI development cohort was insufficient for overall EUS-AI training, and 2) EUS images of the external validation cohort were likely to have different visual data that the EUS-AI system had not experienced during the initial EUS-AI learning process. Notably, based on comparisons of AI and human diagnostic abilities in the external validation cohort, the EUS-AI system was superior to EUS endoscopists in the diagnosis of neoplastic GB polyps and adenocarcinomas. Considering proficiency and EUS experience, the accuracy of the EUS-AI system was between the accuracies of mid-level and expert EUS endoscopists for the differential diagnosis of neoplastic and non-neoplastic polyps. Thus, the accuracy of EUS-AI system may correspond to that of a EUS endoscopist with at least 5 years of experience. If the EUS-AI system is continuously provided with additional training EUS images, its diagnostic performance will further improve.

In clinical practice, previous studies have supported the risks for neoplastic change, which have been used to assess preoperative diagnosis.^{1,2,4,6,15,16} Polyp size is the best predictor of malignancy; several other features on EUS images are used for preoperative diagnosis.

However, excluding the size criterion, these findings on EUS images showed inadequate performance because many parameters in GB polyps are subjective (e.g., hyperechoic spots, microcysts, central hypoechoic foci, and Doppler flow). In addition, the interpretation of EUS and contrast enhanced harmonic-EUS images by visual assessment is subjective; an objective analytical technique for discrimination of GB diseases is required. The EUS-AI system can objectively determine the neoplastic risk of GB polyps before cholecystectomy via intuitive imaging analysis. Moreover, because the size of the polyp is stored in EUS images by the placement of size information into the EUS pixel, connected with the EUS bar ruler, EUS-AI can complete the diagnostic process with consideration of lesion size. Thus, the AUROC of the EUS-AI system for differentiation of neoplastic and non-neoplastic GB polyps in the external validation cohort was 0.69 (95% CI 0.68–0.70), which was higher than the AUROC (0.60 [0.46–0.74]) of the size criterion > 10 mm (standard cut-off value for cholecystectomy).

This study is the first to apply AI for the EUS diagnosis of GB disease. AI has been used in many medical fields such as ophthalmology, dermatology, and oncology.¹⁷⁻²⁰ Recently, an AI diagnostic algorithm using endoscopic images has been developed for gastrointestinal diseases such as peptic ulcers, colorectal polyps, and inflammatory bowel disease.^{11,12,21-26} Regarding AI for EUS images, a few reports have shown the usefulness of AI for pancreatic neoplasms.²⁷⁻²⁹ Similar to our AI analysis tools, Kuwahara et al. used a convolutional neural network (CNN) architecture and demonstrated the superiority of the AI algorithm for the diagnosis of malignant intraductal papillary mucinous neoplasms.²⁹

The current study indicated that the EUS-AI system is useful for objective diagnosis of neoplastic GB polyps; however, there were several limitations. First, the retrospective design and limited sample size have inherent limitations. We used data augmentation to compensate

for the lack of AI training data, but this approach is somewhat less suitable for AI training. To minimize artificial changes in the original image, we used a rigid transformation method for data augmentation. However, data augmentation is considered a temporary measure to compensate for a small sample size; greater numbers of patients are required in future studies for a reliable AI model. Second, because our analysis was limited to representative EUS images, potential EUS findings may have been exaggerated or missed. Because the EUS test observes target lesions in real time, future AI studies are required to assess moving images. Lastly, the current study only included GB polyps with a maximum diameter of 7–20 mm for the AI analysis. Laparoscopic cholecystectomy is considered a simple and standard treatment strategy for polyps >10 mm. Thus, there is a fundamental concern regarding whether additional EUS inspection is required. However, even for 10–20 mm GB polyps, since many comprise both non-neoplastic GB polyps and small gallstones, EUS AI analysis is expected to reduce unnecessary surgeries.

In conclusion, the newly developed EUS-AI system showed favorable performance for diagnosis of neoplastic GB polyps and adenocarcinomas, with performance comparable to that of EUS endoscopists. Because the application of AI has been slow in the EUS field compared with other endoscopic methods, additional efforts are needed to overcome the limitations of EUS, such as the subjective interpretation and lack of standardization. It is expected that the applications of AI in the EUS field will increase with further prospective multicenter studies and advances in technology that enable analysis of EUS videos in real time.

Patients and methods

Preoperative EUS images of pathologically confirmed GB polyps and gallstones were retrospectively collected between January 2014 and May 2019 as the AI development cohort. The Institutional Review Board of Gangnam Severance Hospital, South Korea, approved this study (IRB No. 3-2020-0089). The study was performed in accordance with relevant guidelines/regulations, and informed consent was obtained from all participants and/or their legal guardians. All EUS examinations were performed with a real-time, gray-scale sector scan echoendoscope EUS (GF-UCT260 or GF-UE260-AL5; Olympus, EG3870UTK; Pentax, Tokyo, Japan) by well-trained endoscopists at the EUS unit of Gachon University Gil Medical Center and Yonsei University College of Medicine, Gangnam Severance Hospital. The inclusion criterion was the presence of GB polyps with maximum diameter of 7–20 mm. The AI development cohort consisted of 1,039 EUS images of GB lesions (836 GB polyps and 203 gallstones) from 670 patients who underwent cholecystectomy. An external validation cohort was created with 83 typical EUS images from 83 patients who underwent preoperative EUS and cholecystectomy between June 2019 and January 2020, which allowed additional verification of the EUS-AI algorithm. The diagnostic performance of GB polyps using the EUS-AI system was compared with the diagnostic performances of seven EUS endoscopists; interobserver agreement was determined between the endoscopist groups and AI. The endoscopists were divided into groups according to proficiency and experience with EUS: < 6 months of experience, junior (n = 2); 6 months to 5 years of experience, mid-level (n = 3); and \geq 5 years of experience, expert (n = 2). Institutional review board approval was obtained for this study protocol (Gil Medical Center IRB No. GCIRB2020-126, Gangnam Severance Hospital IRB No. 3-2020-0085).

Data preparation (training and test data)

The EUS data of AI development cohort were divided into training, validation, and test datasets, and the lack of data was supplemented through data augmentation. Data augmentation is defined as an algorithm that artificially expands the data when the training set data are insufficient.³⁰ Data augmentation was accomplished by transforming the original images into arbitrary combinations for flip, rotation, scale, and transformation, respectively. Data augmentation was performed only when insufficient data were available to match the ratio when the number of data deviations was large; the ratio between the two clusters was augmented 5–10-fold if the training data were < 100 images. Data augmentation was applied to the training and validation sets, but not the test set (Supplementary Table 1). Additionally, a separate set of cohort validation data was used for external validation of each model learned. In total, 83 EUS images were collected: 11 gallstones and 72 GB polyps (21 neoplastic, 51 non-neoplastic, 12 adenomatous, and 9 adenocarcinoma). No separate data enhancement was applied to the data set of external validation cohort.

Effective learning data was organized by preventing the inclusion of unnecessary areas through a pre-processing process. The region of interest (ROI) in the form of a box (20 mm × 20 mm) containing lesions was extracted from the EUS image and used in the experiment. Images that were cut to the same size were resampled to the same pixel length (256 × 256 pixels) to convert them into deep learning training data.

Convolutional Neural Network Model for deep learning

The ResNet50 architecture, a convolutional neural network for deep learning, was used in this study. The ResNet50 is a structure in which the two-layer block of the existing 34-layer net has been replaced with a three-layer bottleneck block, thereby maximizing image classification

performance due to the deep network layer and the low error rate.^{31,32} The hyperparameter for training was fixed at 40 for batch size, 300 for epochs, and 0.0001 for the learning rate; this was applied equally to all discriminative models (Fig. 1). To understand and interpret the classification results of the trained model, the class activation map was obtained by visualizing the feature map of the last layer in the trained model(Supplementary Fig 1).^{33,34}

Statistical analysis

Three types of classification models were used for the EUS-AI system (model for GB polyp vs. gallstones, model for neoplastic polyps vs. non-neoplastic polyps, and model for adenocarcinoma vs. adenomatous GB polyps); these models were trained using the deep learning algorithm. Each trained model was evaluated for performance through five cross-validations for robustness and data dependence.³⁵ The models trained from each cross-validation produced true positives, true negatives, false positives, and false negatives according to the classification results; these data were used to calculate statistical measures of sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy. The area under the receiver operating characteristic curve (AUROC) was also calculated. Interobserver agreement was calculated using Cohen's kappa or Fleiss' kappa. Statistical analysis of clinical and demographic data was performed using SPSS software (version 23.0, IBM Corp., Armonk, NY, USA); $p < 0.05$ was considered statistically significant.

Reference

1. Sugiyama, M., Atomi, Y. & Yamato, T. Endoscopic ultrasonography for differential diagnosis of polypoid gall bladder lesions: analysis in surgical and follow up series. *Gut* **46**, 250-254 (2000).
2. Cho, J. H. *et al.* Hypoechoic foci on EUS are simple and strong predictive factors for neoplastic gallbladder polyps. *Gastrointestinal endoscopy* **69**, 1244-1250 (2009).
3. Fujita, N., Noda, Y., Kobayashi, G., Kimura, K. & Yago, A. Diagnosis of the depth of invasion of gallbladder carcinoma by EUS. *Gastrointestinal endoscopy* **50**, 659-663 (1999).
4. Kim, S. Y. *et al.* The efficacy of real-time colour Doppler flow imaging on endoscopic ultrasonography for differential diagnosis between neoplastic and non-neoplastic gallbladder polyps. *European radiology* **28**, 1994-2002 (2018).
5. Sadamoto, Y. *et al.* A useful approach to the differential diagnosis of small polypoid lesions of the gallbladder, utilizing an endoscopic ultrasound scoring system. *Endoscopy* **34**, 959-965 (2002).
6. Choi, W. B. *et al.* A new strategy to predict the neoplastic polyps of the gallbladder based on a scoring system using EUS. *Gastrointestinal endoscopy* **52**, 372-379 (2000).
7. Park, C. H. *et al.* Differential diagnosis between gallbladder adenomas and cholesterol polyps on contrast-enhanced harmonic endoscopic ultrasonography. *Surgical endoscopy* **27**, 1414-1421 (2013).
8. Choi, J. H. *et al.* Utility of contrast-enhanced harmonic EUS in the diagnosis of malignant gallbladder polyps (with videos). *Gastrointestinal endoscopy* **78**, 484-493 (2013).

9. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436-444 (2015).
10. Mori, Y. *et al.* Cost savings in colonoscopy with artificial intelligence-aided polyp diagnosis: an add-on analysis of a clinical trial (with video). *Gastrointestinal endoscopy* (2020).
11. Sanchez-Montes, C. *et al.* Computer-aided prediction of polyp histology on white light colonoscopy using surface pattern analysis. *Endoscopy* **51**, 261-265 (2019).
12. Wu, L. L. *et al.* A deep neural network improves endoscopic detection of early gastric cancer without blind spots. *Endoscopy* **51**, 522-531 (2019).
13. Wiles, R. *et al.* Management and follow-up of gallbladder polyps : Joint guidelines between the European Society of Gastrointestinal and Abdominal Radiology (ESGAR), European Association for Endoscopic Surgery and other Interventional Techniques (EAES), International Society of Digestive Surgery - European Federation (EFISDS) and European Society of Gastrointestinal Endoscopy (ESGE). *European radiology* **27**, 3856-3866 (2017).
14. Wennmacker, S. Z. *et al.* Polyp size of 1 cm is insufficient to discriminate neoplastic and non-neoplastic gallbladder polyps. *Surgical endoscopy* **33**, 1564-1571 (2019).
15. Sugiyama, M., Xie, X. Y., Atomi, Y. & Saito, M. Differential diagnosis of small polypoid lesions of the gallbladder: the value of endoscopic ultrasonography. *Ann Surg* **229**, 498-504 (1999).
16. Azuma, T., Yoshikawa, T., Araida, T. & Takasaki, K. Differential diagnosis of polypoid lesions of the gallbladder by endoscopic ultrasonography. *American journal of surgery* **181**, 65-70 (2001).

17. Ting, D. S. W. *et al.* Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes. *JAMA* **318**, 2211-2223 (2017).
18. Gulshan, V. *et al.* Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA* **316**, 2402-2410 (2016).
19. Esteva, A. *et al.* Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542**, 115-118 (2017).
20. Ehteshami Bejnordi, B. *et al.* Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *JAMA* **318**, 2199-2210 (2017).
21. Lee, J. H. *et al.* Spotting malignancies from gastric endoscopic images using deep learning. *Surgical endoscopy* **33**, 3790-3797 (2019).
22. Wang, Z. *et al.* Deep learning-based endoscopic image recognition for detection of early gastric cancer: a Chinese perspective. *Gastrointestinal endoscopy* **88**, 198-199 (2018).
23. Hashimoto, R. *et al.* Artificial intelligence using convolutional neural networks for real-time detection of early esophageal neoplasia in Barrett's esophagus (with video). *Gastrointestinal endoscopy* (2020).
24. Takenaka, K. *et al.* Development and Validation of a Deep Neural Network for Accurate Evaluation of Endoscopic Images From Patients with Ulcerative Colitis. *Gastroenterology* (2020).
25. Mori, Y., Kudo, S. E., Berzin, T. M., Misawa, M. & Takeda, K. Computer-aided diagnosis for colonoscopy. *Endoscopy* **49**, 813-819 (2017).

26. Ichimasa, K. *et al.* Artificial intelligence may help in predicting the need for additional surgery after endoscopic resection of T1 colorectal cancer. *Endoscopy* **50**, 230-240 (2018).
27. Saftoiu, A. *et al.* Efficacy of an artificial neural network-based approach to endoscopic ultrasound elastography in diagnosis of focal pancreatic masses. *Clin Gastroenterol Hepatol* **10**, 84-90 e81 (2012).
28. Saftoiu, A. *et al.* Quantitative contrast-enhanced harmonic EUS in differential diagnosis of focal pancreatic masses (with videos). *Gastrointestinal endoscopy* **82**, 59-69 (2015).
29. Kuwahara, T. *et al.* Usefulness of Deep Learning Analysis for the Diagnosis of Malignancy in Intraductal Papillary Mucinous Neoplasms of the Pancreas. *Clin Transl Gastroenterol* **10**, 1-8 (2019).
30. Taylor, L. & Nitschke, G. Improving deep learning using generic data augmentation. *arXiv preprint arXiv:1708.06020* (2017).
31. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition in *Proceedings of the IEEE conference on computer vision and pattern recognition* 770-778(2016).
32. Canziani, A., Paszke, A. & Culurciello, E. An analysis of deep neural network models for practical applications. *arXiv preprint arXiv:1605.07678* (2016).
33. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A. & Torralba, A. Learning deep features for discriminative localization in *Proceedings of the IEEE conference on computer vision and pattern recognition* 2921-2929(2016).
34. Selvaraju, R. R. *et al.* Grad-cam: Visual explanations from deep networks via gradient-based localization in *Proceedings of the IEEE international conference on computer vision* 618-626(2017).

35. Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection in *Ijcai*, Vol. 14 1137-1145 (Montreal, Canada, 1995).

Acknowledgments: This work was supported by faculty research grant from the Gachon University Gil Medical Center (Kim Eui Joo, Grant number: FRD2019-04) and faculty research grant from Yonsei University College of Medicine (Cho Jae Hee, 6-2020-0133)

Authors Contributions:

S.I.J.: drafting of the article; technical and material support; critical revision of the article for important intellectual content; analysis and interpretation of data

Y.J.K.: drafting of the article; technical and material support; critical revision of the article for important intellectual content; analysis and interpretation of data

E.J.K.: technical and material support; analysis and interpretation of the data

H.K.: technical and material support; analysis and interpretation of the data

S.J.S.: technical and material support; analysis and interpretation of the data

Y.J.S.: technical and material support; analysis and interpretation of the data

D.K.L.: analysis and interpretation of the data; critical revision of the article for important intellectual content

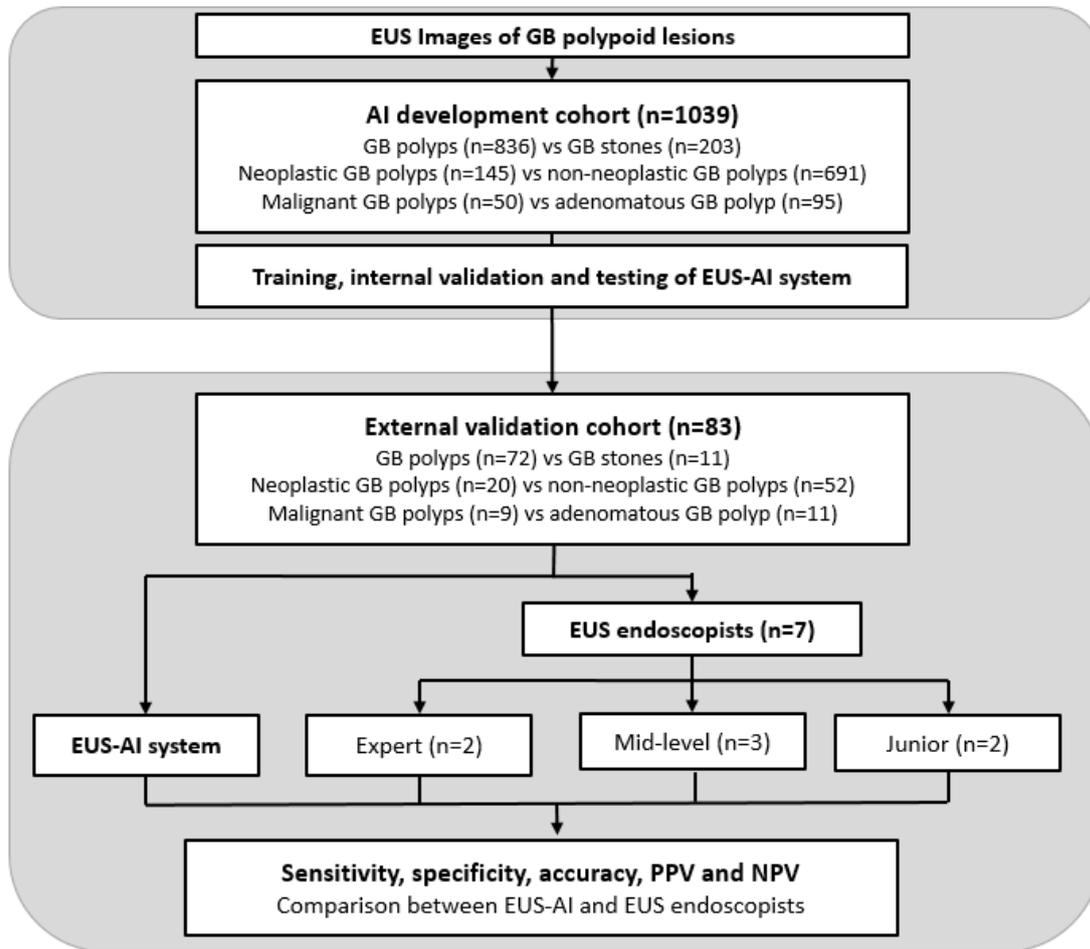
K.G.K.: conception and design; case collection; critical revision of the article for important intellectual content; final approval of the article

J.H.C.: conception and design; case collection; critical revision of the article for important intellectual content; final approval of the article

Conflict of interest: The authors have no conflicts of interest to report

Figure legends

Figure 1. Patient flowchart



GB, gallbladder; EUS, endoscopic ultrasound; AI, artificial intelligence; PPV, positive predictive value; NPV, negative predictive value

Figure 2. Class Activation Map (CAM) analyzes which part of the image affected the classification in the image classification through deep learning.

This figure represents the CAM of the EUS image classified as correct (1,2) and the image classified as incorrect (3,4) in each experiment. A, gallstone. B, gallbladder polyp. C, non-neoplastic gallbladder polyp. D, neoplastic gallbladder polyp

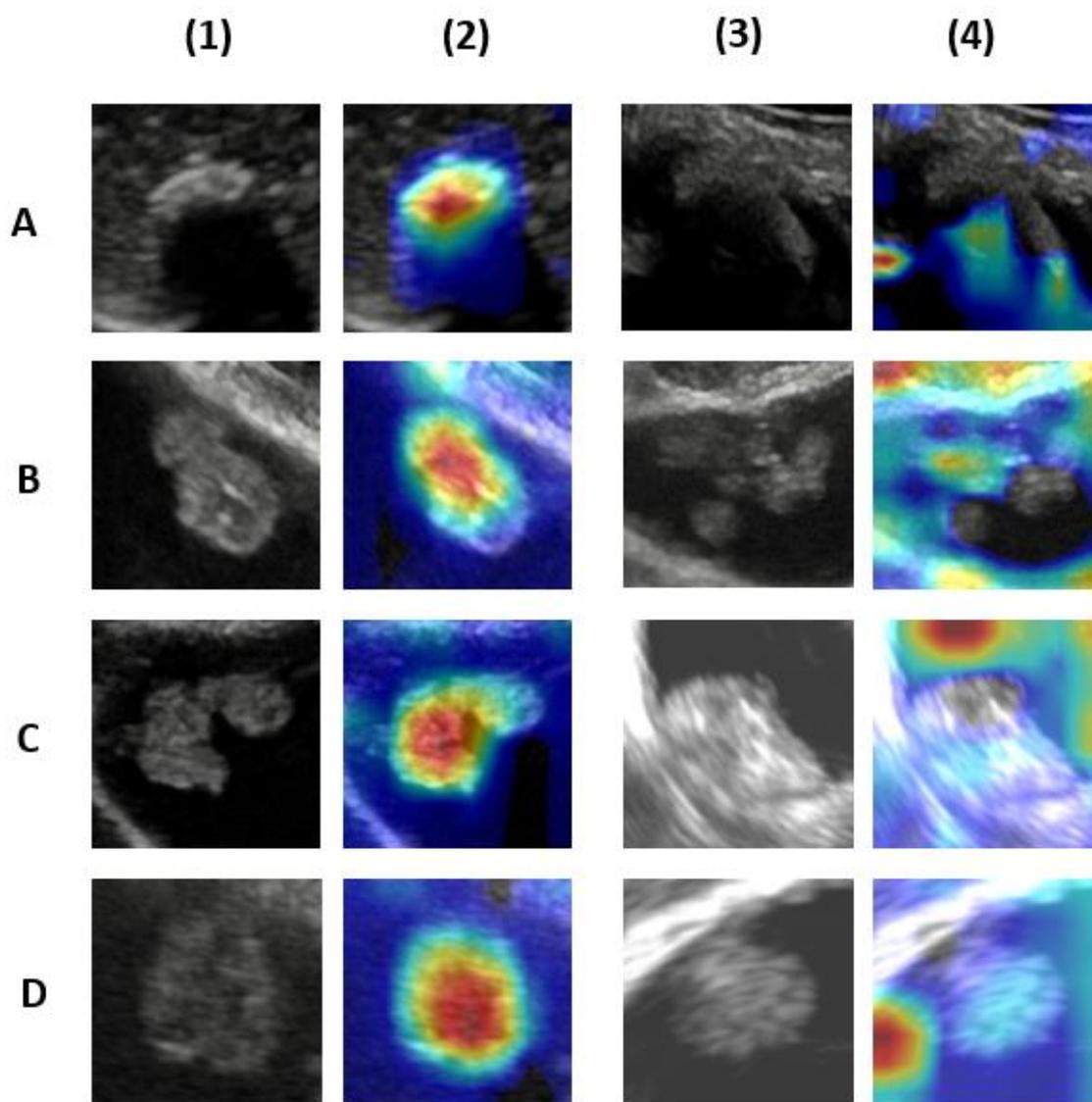
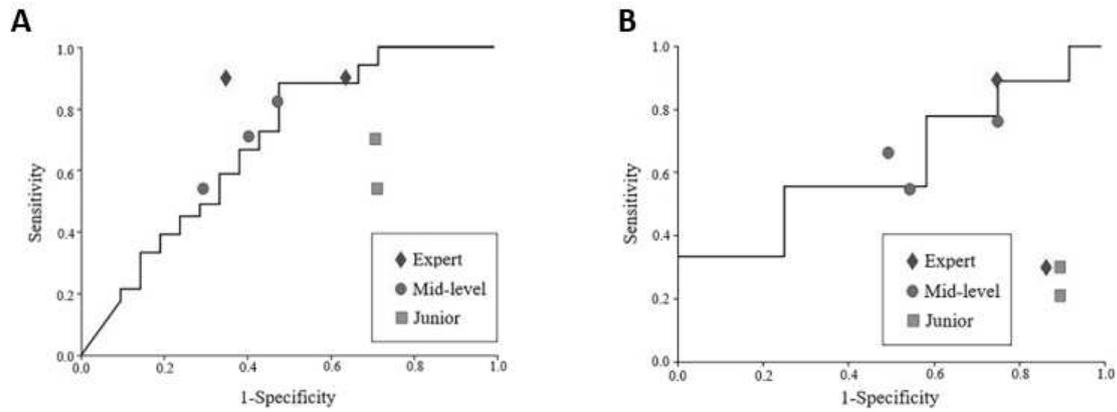


Figure 3. Receiver operating characteristic curve (ROC) for the image analysis based on external validation datasets (n=83).



A, ROC curve for differentiation between neoplastic and non-neoplastic GB polyps. B, ROC curve for differentiation between adenocarcinoma and adenomatous GB polyps

Table 1. Characteristics of patient and gallbladder lesions in the AI development cohort and external validation cohort

Characteristics	GB polyp			Gallstone
	All	Neoplastic polyp	Non-neoplastic polyp	
AI development cohort				
Number of patients (images)	670 (1039)	99 (145)	368 (691)	203 (203)
Patients characteristics				
Age	53.3 ± 13.8	56.5 ± 13.9	51.2 ± 12.1	55.3 ± 16.0
Sex (M/F)	326/344	45/54	192/176	89/114
Lesion characteristics (images)				
Size (mean ± SD)	11.4 ± 3.9	15.1 ± 4.2	10.6 ± 3.1	11.1 ± 4.0
Cholesterol polyp / adenomyomatosis / etc.			638/41/12	
Adenocarcinoma / adenomatous polyps		50/95		
External validation cohort				
Number of patients (images)	83 (83)	20 (20)	52 (52)	11 (11)
Patients characteristics				
Age	53.3 ± 13.3	61 ± 12.5	49.8 ± 11.8	54.9 ± 15.9
Sex (M/F)	37/46	12/8	18/34	5/6
Lesion characteristics (images)				
Size (mean ± SD)	11.8 ± 4.2	15.0 ± 4.1	10.9 ± 3.6	10.2 ± 5.9
Cholesterol polyp / adenomyomatosis / etc.			45/5/2	
Adenocarcinoma / adenomatous polyps		9/11		

GB, gallbladder; M, male; F, female, SD, standard deviation

Table 2. Diagnostic performance of EUS-AI system using train, validation, test set of the AI development cohort according to various diagnostic steps.

		Train	Validation	Test
Step 1 Differential diagnosis between GB polyps and gallstones	Sensitivity (%)	99.3	99.1	96.3
	Specificity (%)	100	100	92.9
	PPV (%)	100	100	98.4
	NPV (%)	99.3	99.1	79.5
	Accuracy (%)	99.7	99.5	95.7
	AUC (95% CI)	-	-	0.990 (0.984-0.994)
Step 2 Differential diagnosis between neoplastic and non-neoplastic GB polyps	Sensitivity (%)	98.2	98.5	77.8
	Specificity (%)	100	100	91.6
	PPV (%)	100	100	57.9
	NPV (%)	98.1	98.4	96.5
	Accuracy (%)	99.1	99.2	89.8
	AUC (95% CI)	-	-	0.891 (0.868-0.913)-
Step 3	Sensitivity (%)	100	100	85.6

Differential diagnosis between adenocarcinoma and adenomatous GB polyps	Specificity (%)	100	100	75.0
	PPV (%)	100	100	87.4
	NPV (%)	100	100	72.0
	Accuracy (%)	100	100	82.1
	AUC (95% CI)	-	-	0.910 (0.819-0.999)-

EUS, endoscopic ultrasound; AI, artificial intelligence; GB, gallbladder; PPV, Positive predictive value; NPV, Negative predictive value; CI, confidence interval

Table 3. Diagnostic performance of EUS-AI system versus endoscopists in differential diagnosis of GB polyps using the external validation cohort (n=83).

	GB-AI system	Endoscopists				
		All	Expert	Mid-level	Junior	
Step 1 Differential diagnosis between GB polyps and gallstones	Sensitivity (%)	98.4	74.7	90.9	69.7	63.6
	Specificity (%)	45.8	97.9	97.9	100	95.8
	PPV (%)	83.6	86.1	87.5	100	70.7
	NPV (%)	90.9	96.2	98.6	95.6	94.5
	Accuracy (%)	84.6	94.9	97	96	91.6
	AUC (95% CI)	0.948 (0.932-0.964)				
Step 2 Differential diagnosis between neoplastic and non-neoplastic GB polyps	Sensitivity (%)	60.3	74.2	90.2	69.9	62.6
	Specificity (%)	77.4	44.9	47.6	58.7	28.3
	PPV (%)	36.2	75.4	80.9	80.4	64.9
	NPV (%)	90.2	46.2	65.2	46.3	27
	Accuracy (%)	74.4	65.3	77.5	66.7	51.4
	AUC (95% CI)	0.691 (0.679-0.702)				
Step 3 Differential diagnosis between adenocarcinoma and adenomatous GB polyps	Sensitivity (%)	65.3	51.8	61.1	66.7	27.8
	Specificity (%)	60.6	21.8	20.9	36.1	8.3
	PPV (%)	78.3	32.6	35.1	44.3	18.4
	NPV (%)	44.4	40.3	50	57.4	13.4
	Accuracy (%)	63.8	34.7	38.1	49.2	16.7
	AUC (95% CI)	0.634 (0.569-0.700)				

EUS, endoscopic ultrasound; AI, artificial intelligence; GB, gallbladder; PPV, Positive predictive value; NPV, Negative predictive value; CI, confidence interval

Figures

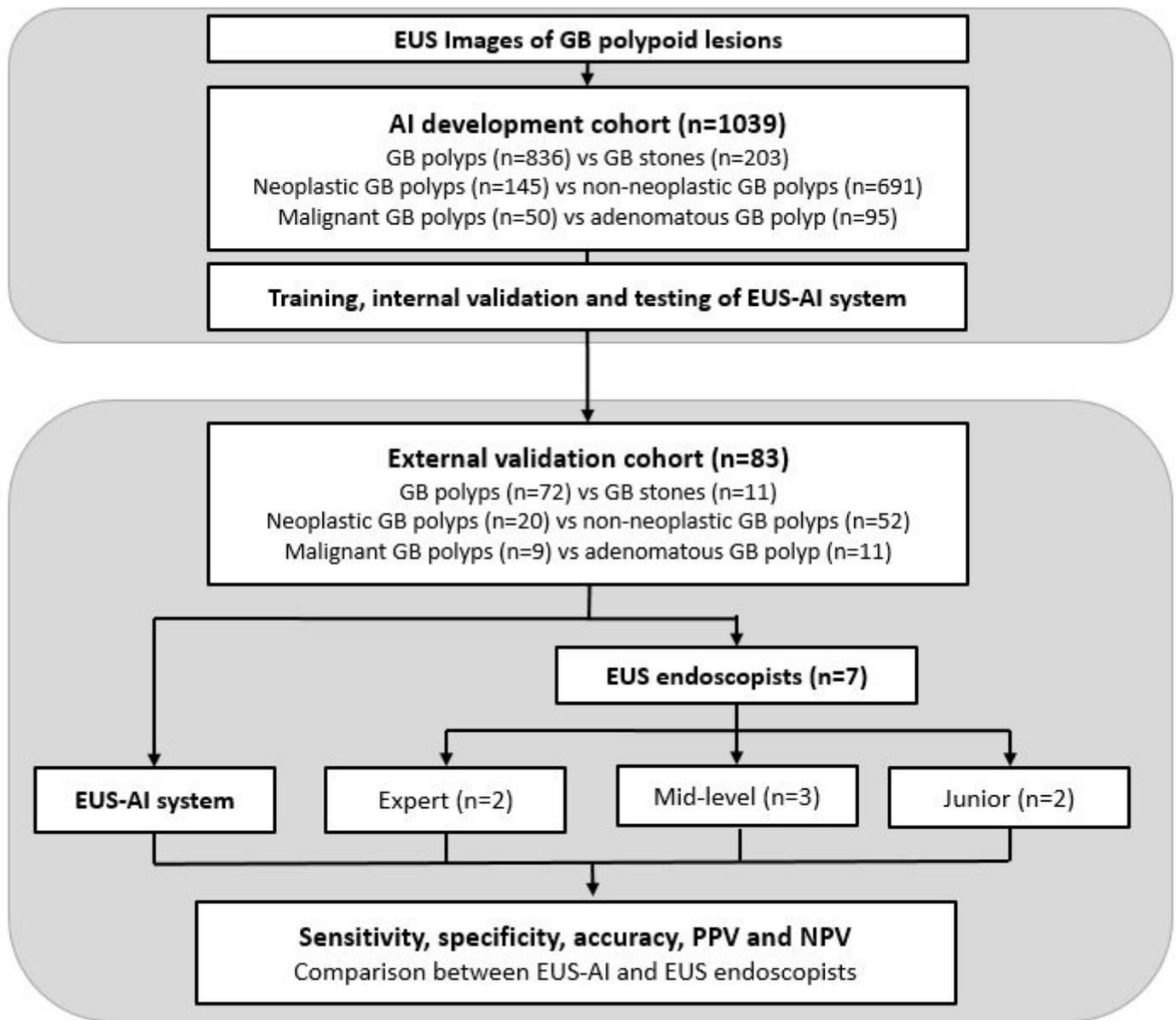


Figure 1

Patient flowchart. GB, gallbladder; EUS, endoscopic ultrasound; AI, artificial intelligence; PPV, positive predictive value; NPV, negative predictive value

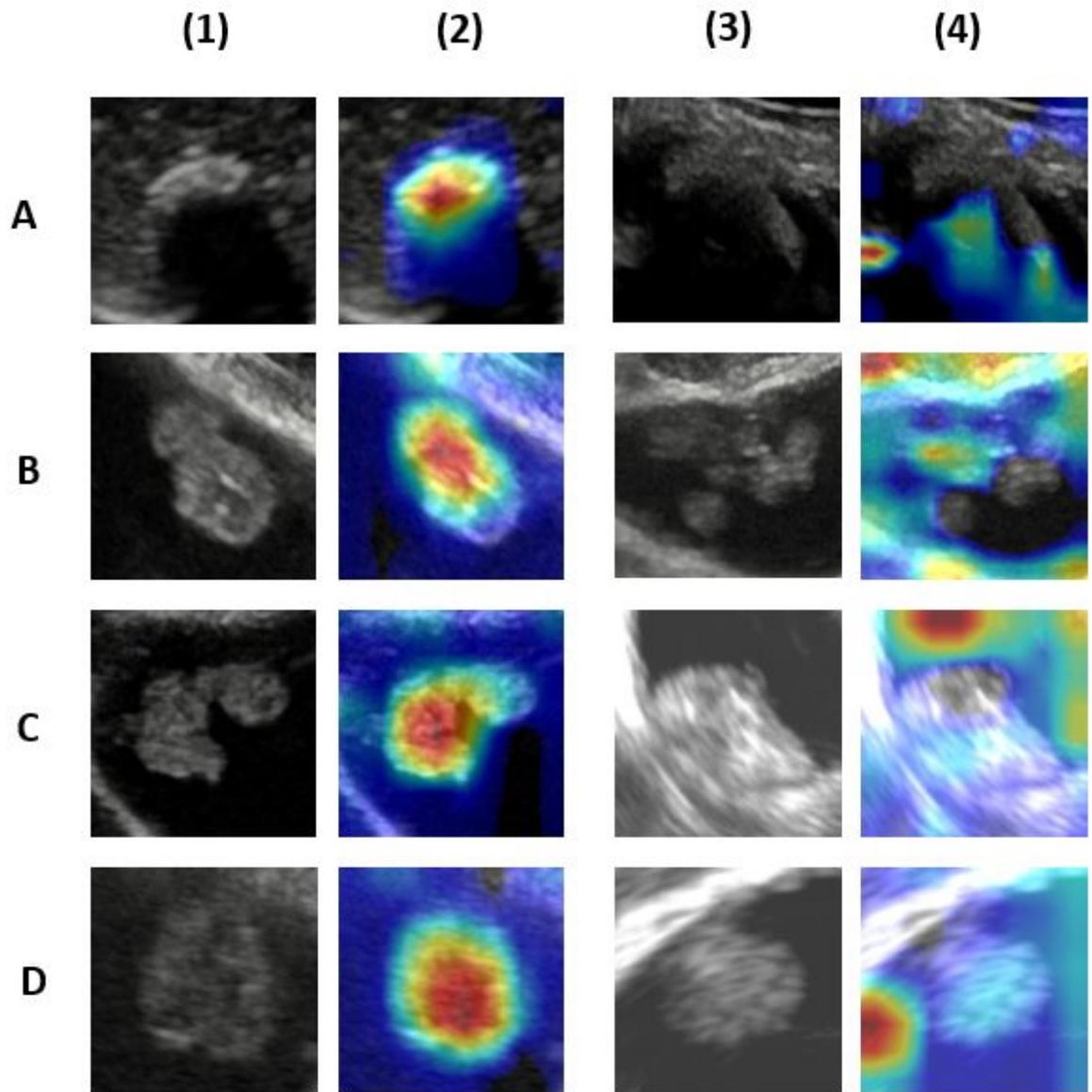


Figure 2

Class Activation Map (CAM) analyzes which part of the image affected the classification in the image classification through deep learning. This figure represents the CAM of the EUS image classified as correct (1,2) and the image classified as incorrect (3,4) in each experiment. A, gallstone. B, gallbladder polyp. C, non-neoplastic gallbladder polyp. D, neoplastic gallbladder polyp

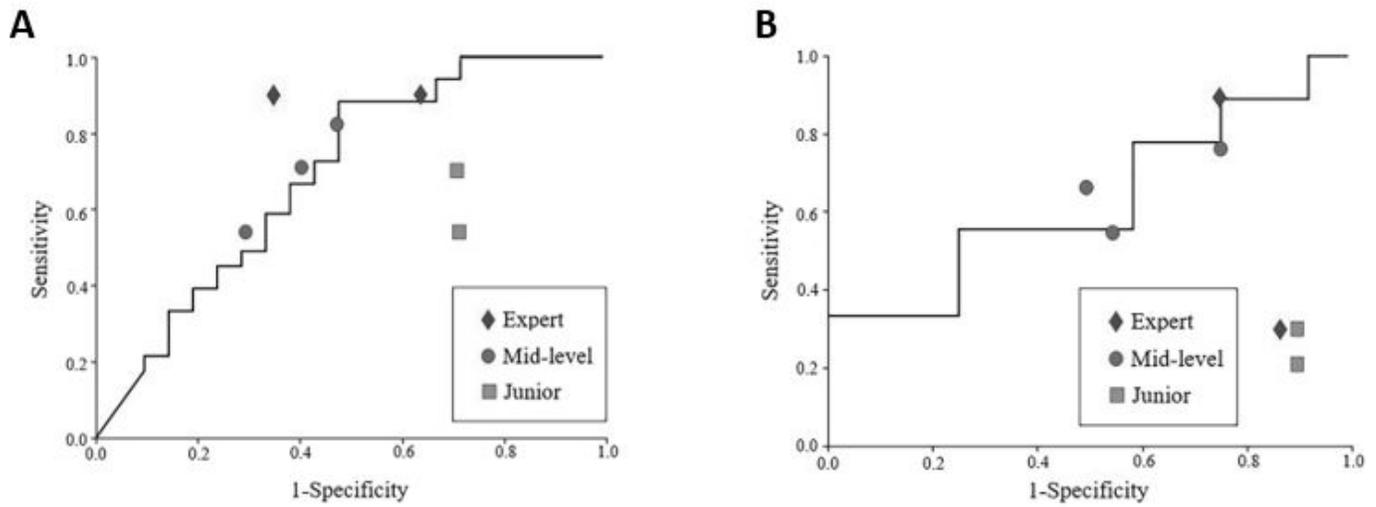


Figure 3

Receiver operating characteristic curve (ROC) for the image analysis based on external validation datasets (n=83). A, ROC curve for differentiation between neoplastic and non-neoplastic GB polyps. B, ROC curve for differentiation between adenocarcinoma and adenomatous GB polyps

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementaryinformation.pdf](#)