

# Artificial Intelligence in the Prediction of Gastrointestinal Stromal Tumors on Endoscopic Ultrasonography Images: Development, Validation and Comparison with Endosonographers

Yi Lu<sup>1,2</sup>, Jiachuan Wu<sup>3</sup>, Minhui Hu<sup>1,2</sup>, Qinghua Zhong<sup>4</sup>, Limian Er<sup>5</sup>, Huihui Shi<sup>5</sup>, Weihui Cheng<sup>6</sup>, Ke Chen<sup>7</sup>, Yuan Liu<sup>7</sup>, Bingfeng Qiu<sup>8</sup>, Qiancheng Xu<sup>8</sup>, Guangshun Lai<sup>9</sup>, Yufeng Wang<sup>10</sup>, Yuxuan Luo<sup>10</sup>, Jinbao Mu<sup>10</sup>, Wenjie Zhang<sup>11</sup>, Min Zhi<sup>2,12</sup>, and Jiachen Sun<sup>1,2</sup>

<sup>1</sup>Department of Gastrointestinal Endoscopy and <sup>2</sup>Guangdong Provincial Key Laboratory of Colorectal and Pelvic Floor Diseases, The Sixth Affiliated Hospital, Sun Yat-sen University, <sup>3</sup>Digestive Endoscopy Center, Guangdong Second Provincial General Hospital,

<sup>4</sup>Department of Endoscopic Surgery, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, <sup>5</sup>Department of Endoscopy, The Fourth Hospital of Hebei Medical University, Shijiazhuang, <sup>6</sup>Department of Gastroenterology, Yangjiang Hospital of Traditional Chinese Medicine, Yangjiang, <sup>7</sup>Department of Endoscopy, Fudan University Shanghai Cancer Center, Shanghai, <sup>8</sup>Department of Gastroenterology, Zhoushan Hospital of Zhejiang Province, Zhoushan, <sup>9</sup>Department of Gastroenterology, Lianjiang People's Hospital, Lianjiang, <sup>10</sup>Tianjin Economic-Technological Development Area (TEDA) Yujin Digestive Health Industry Research Institute, <sup>11</sup>Tianjin Center for Medical Devices Evaluation and Inspection, Tianjin, and <sup>12</sup>Department of Gastroenterology, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, China

## Article Info

Received August 8, 2022

Revised September 19, 2022

Accepted October 7, 2022

Published online January 26, 2023

## Corresponding Author

Min Zhi

ORCID <https://orcid.org/0000-0001-8178-5572>

E-mail zhimin@mail.sysu.edu.cn

Jiachen Sun

ORCID <https://orcid.org/0000-0003-3646-1039>

E-mail sunjch8@mail.sysu.edu.cn

Yi Lu and Jiachuan Wu contributed equally to this work as first authors.

**Background/Aims:** The accuracy of endosonographers in diagnosing gastric subepithelial lesions (SELs) using endoscopic ultrasonography (EUS) is influenced by experience and subjectivity. Artificial intelligence (AI) has achieved remarkable development in this field. This study aimed to develop an AI-based EUS diagnostic model for the diagnosis of SELs, and evaluated its efficacy with external validation.

**Methods:** We developed the EUS-AI model with ResNeSt50 using EUS images from two hospitals to predict the histopathology of the gastric SELs originating from muscularis propria. The diagnostic performance of the model was also validated using EUS images obtained from four other hospitals.

**Results:** A total of 2,057 images from 367 patients (375 SELs) were chosen to build the models, and 914 images from 106 patients (108 SELs) were chosen for external validation. The sensitivity, specificity, positive predictive value, negative predictive value, and accuracy of the model for differentiating gastrointestinal stromal tumors (GISTs) and non-GISTs in the external validation sets by images were 82.01%, 68.22%, 86.77%, 59.86%, and 78.12%, respectively. The sensitivity, specificity, positive predictive value, negative predictive value, and accuracy in the external validation set by tumors were 83.75%, 71.43%, 89.33%, 60.61%, and 80.56%, respectively. The EUS-AI model showed better performance (especially specificity) than some endosonographers. The model helped improve the sensitivity, specificity, and accuracy of certain endosonographers.

**Conclusions:** We developed an EUS-AI model to classify gastric SELs originating from muscularis propria into GISTs and non-GISTs with good accuracy. The model may help improve the diagnostic performance of endosonographers. Further work is required to develop a multi-modal EUS-AI system. (*Gut Liver* 2023;17:874-883)

**Key Words:** Artificial intelligence; Subepithelial lesions; Gastrointestinal stromal tumors; Endoscopic ultrasonography; Gastric

## INTRODUCTION

The reported detection rates of gastric subepithelial

lesions (SELs) during upper gastrointestinal endoscopy ranged from 0.3% to 2%.<sup>1-3</sup> Among these, gastrointestinal stromal tumors (GISTs) and leiomyomas are more com-

mon than other types.<sup>4</sup> GISTs have malignant potential, and need to be cautiously monitored or resected.<sup>4,5</sup> The lesions are usually located at the muscularis propria.<sup>6</sup> Accurate differentiation of GISTs and non-GISTs is of much clinical importance. Endoscopic ultrasonography (EUS) is usually recommended as the subsequent examination modality.<sup>7</sup> However, the EUS-based diagnosis of SELs often depends on the experience of the operator and is liable to be influenced by subjective factors. The diagnosis is usually made based on the analysis of its location, originating layer, margins, echogenicity, and morphology.

The reported accuracy rate of the endoscopists in the diagnosis of SELs using EUS ranged from 43% to 73%.<sup>6,8-10</sup> In order to improve the diagnostic accuracy, some experts have used artificial intelligence (AI) systems to develop convolutional neural networks models to help predict the diagnosis of SELs on EUS images. Minoda *et al.*<sup>8</sup> developed an EUS diagnostic system with AI (EUS-AI), and demonstrated a good diagnostic yield of their EUS-AI system for SELs  $\geq 20$  mm; however, the accuracy, especially specificity for SELs  $< 20$  mm was not that satisfactory.<sup>8</sup> We believe it is more important to distinguish the smaller SELs, as resection is usually recommended for SELs  $\geq 20$  mm. Ability to predict smaller-sized SELs would help inform treatment decision-making in clinical settings. Hence, in this study, we sought to develop an AI-based EUS diagnostic model for the diagnosis of SELs, and evaluated its efficacy with both internal and external validation.

## MATERIALS AND METHODS

### 1. Clinical data and EUS image collection

EUS images were retrospectively searched at the Sixth Affiliated Hospital, Sun Yat-sen University and Guangdong Second Provincial Central Hospital to build the model. The inclusion criteria for the collected images were: good-quality EUS images showing the tumors; gastric SELs originating from muscularis propria; confirmed histopathology (obtained by endoscopic resection, surgery or fine needle aspiration). The exclusion criteria were: histopathology result not in accordance with the clinical situation (for example, resection of just the mucous membrane rather than the tumor, or negative fine needle aspiration), leiomyosarcoma or other sarcomas, poor quality images, and repeated images. The following information of the patients was also collected: age, sex, EUS results, and histopathology results. The collected images were categorized into two groups based on the histopathology results of the SELs: GISTs and non-GISTs.

### 2. EUS procedures

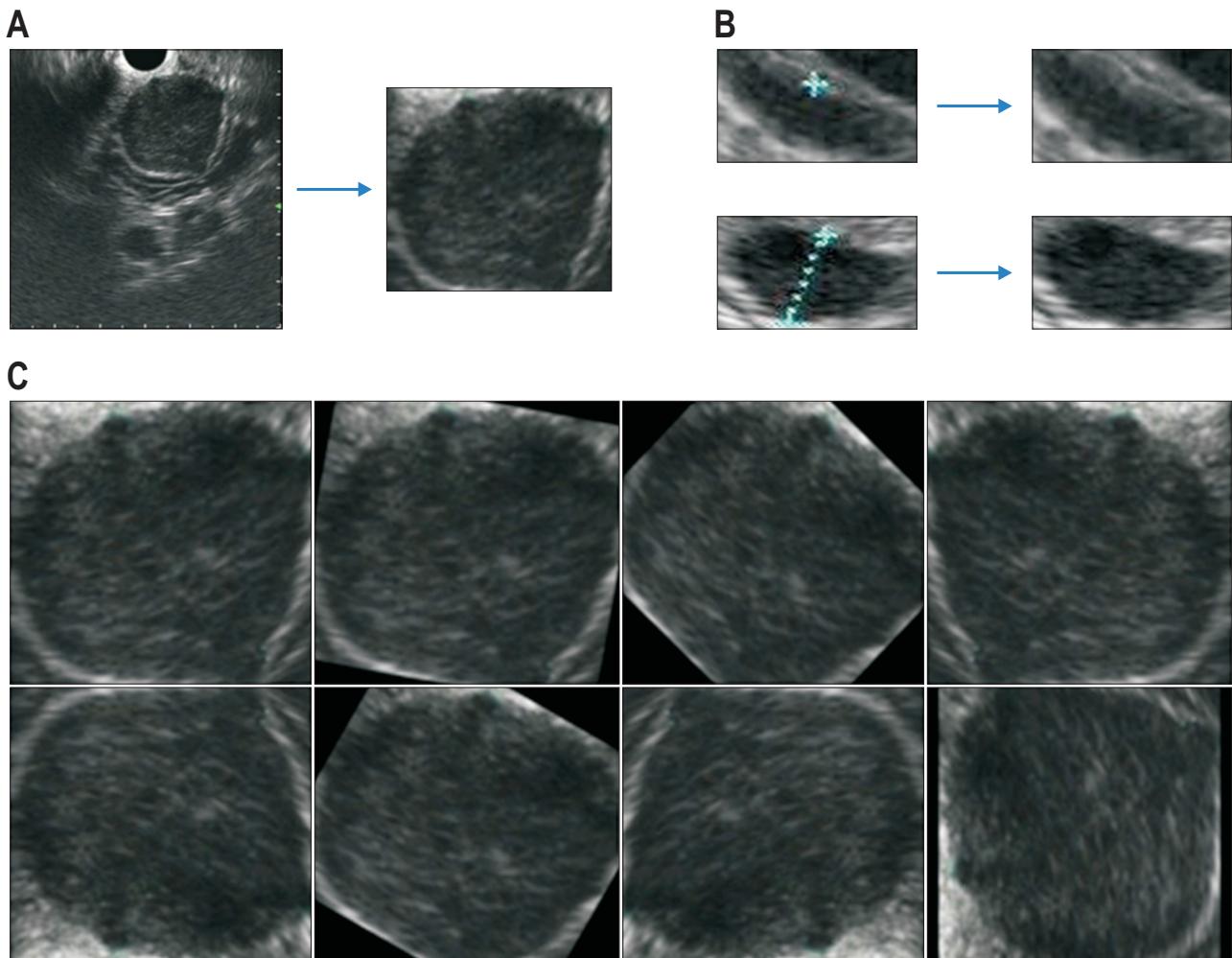
EUS was performed by endosonographers with experience of diagnosing more than 500 cases of SELs by EUS. Conventional machines were used: EU-ME1, EU-ME2 (Olympus, Tokyo, Japan), MAJ-1720 (Olympus), SU-9000 (Fujifilm, Tokyo, Japan), SP-900 (Fujifilm), and HI VISION Preirus (HITACHI, Tokyo, Japan). The echoendoscopes used were GF-UE260-AL5, GF-UCT240-AL5, and GF-UCT260 (Olympus, frequency 5–12 MHz), mini-probes UM-2R or UM-3R (Olympus, frequency 20 MHz), EG-530UT2, EG-580UT, or EG-580UR (Fujifilm, frequency 5–12 MHz), mini-probes P2612-M or P2615-M (Fujifilm, frequency 12 MHz or 15 MHz), and EG-3270UK, EG-3670URK, EG-3830UT or EG-3870UTK (Pentax Lifecare, Tokyo, Japan, frequency 5–20 MHz). The echoendoscopes used in the models and the external validation sets are specified in the Supplementary Table 1, and the echoendoscopes used in each hospital are specified in Supplementary Table 2.

### 3. Image processing and augmentation

After collecting and selecting the qualified images, two experts (Y. Lu and J.S) in EUS firstly marked the border of the tumor with software named LabelMe, which is a polygonal and open annotation tool developed with Python by the Massachusetts Institute of Technology, Computer Science and Artificial Intelligence Laboratory. The marked tumors were regarded as the regions of interest. Then the engineers trimmed the images to squares or rectangles precisely fitting the regions of interest (Fig. 1A). Some images had measuring lines or measuring marks which would have affected the accuracy of the deep learning models; for such images, we used “clone stamp” tool of the Adobe Photoshop (version 13.0) erase them (Fig. 1B), and we only chose those with a little measuring lines or marks so as to preserve the original images, if the change was huge, then we would not choose this image. To enlarge the training sets, image augmentation technology was applied. We used mirror flip, horizontal flip, and rotation in certain degrees without disturbing the textures of the EUS images (Fig. 1C). The preprocessed images were then changed into the RGB three-channel to be regarded as the model input.

### 4. Development of the deep learning models and external validation

We used deep convolutional neural networks classifier named ResNeSt50 to train the model.<sup>11</sup> The image processing, augmentation, and the development of the deep learning models was supported by the Tianjin Jinyu Artificial Intelligence Medical Technology Co., Ltd. The process was mainly performed by Python (version 3.7) and PyTorch



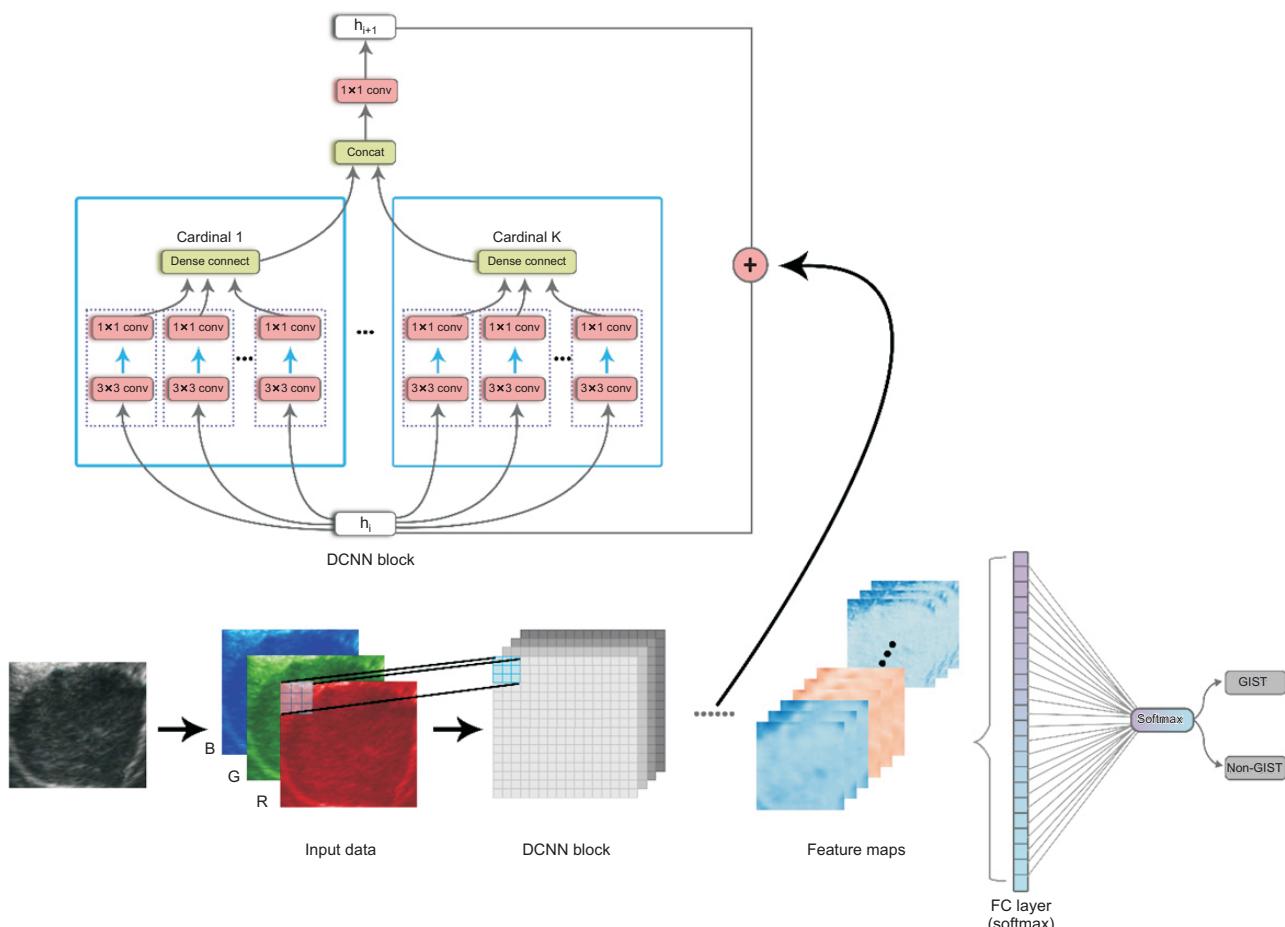
**Fig. 1.** Images process. (A) Images were trimmed to squares or rectangles containing the regions of interest. (B) Erasure of the measuring lines or measuring marks. (C) Images showing mirror flip, horizontal flip, and rotation in certain degrees for image augmentation.

(version 1.7.1). The images chosen were randomly divided into training sets and test sets with a ratio of 9:1, and 10-fold cross-validation was applied. The stochastic gradient descent optimizer was used, and we introduced the first-order momentum (whose value was 0.9), which made the gradient update inertially and could better achieve the effect of convergence. The initial learning rate was set as 1e-3, and cosine annealing was used as the attenuation method. The model was trained for 600 epochs, and the final learning rate was 1e-6. When the training reached the preset loss value (loss <0.0005) or 600 epochs, the training was stopped (Fig. 2 shows the schematic diagram). The output of the model was the probability for the pathological type of the SELs (GIST or non-GIST) based on the evaluation of the EUS images. Three experts (Y. Lu, J.S., and W.C.; experience of more than 1,000 EUS procedures) and three novices (J.W., M.H., and G.L.; experience of less than 500 EUS procedures) who were blinded to the pathology of the SELs independently judged the classification of

the lesions.

Then, we collected the EUS images from four other hospitals (Fudan University Shanghai Cancer Center, the Fourth Hospital of Hebei Medical University, Zhoushan Hospital of Zhejiang Province, and Yangjiang Hospital of Traditional Chinese Medicine) to perform external validation. The inclusion and exclusion criteria were the same as mentioned above for the selection of the images for building the model. Images in the external validation dataset were also marked with regions of interests and the measuring lines and marks were erased.

In the external validation, we also evaluated the diagnostic performance of AI by tumors. Firstly, the number of images of the tumor diagnosed by AI with GISTs or non-GISTs was counted, if more images were diagnosed as GISTs, then the result of this tumor was GISTs, and vice versa. If the number was equal, then the pooled predictive probability was calculated, and the category with larger probability was the final diagnosis. Also, three experts and



**Fig. 2.** Schematic diagram of the DCNNs from an endoscopic ultrasonography image to the final prediction. DCNN, deep convolutional neural networks; FC, fully connected; conv, convolutional layer; GIST, gastrointestinal stromal tumor.

three novices who were blinded to the histopathological results, were asked to classify the SELs in the external validation sets, and then they were asked to classify the SELs again after knowing the diagnosis of the EUS-AI model (the endosonographers were not forced to accept the results of AI models, rather, if they know the results of the AI models, they would think twice and make the final diagnosis by their own choice of whether to believe AI models or believe themselves).

## 5. Statistical analysis

Categorical variables were expressed as number (percentage). Normally distributed continuous variables were presented as mean (standard deviation), while non-normally distributed continuous variables were presented as median (range). Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy of the EUS-AI model and the respective 95% confidence intervals (CIs) were calculated. The receiver operating characteristic curve was plotted, and the area under the curve was calculated. The above calculation was performed

by the Scikit-learn package in Python. The accuracy, sensitivity, and specificity were compared using the chi-square test (SPSS Statistic version 26; IBM Corp., Armonk, NY, USA). Two-tailed p-values  $<0.05$  were considered indicative of statistical significance.

This study was approved by the Institutional Review Board of The Sixth Affiliated Hospital, Sun Yat-sen University (IRB number: 2021ZSLYEC-319), and the informed consent was waived. The study has been registered in the Chinese Clinical Trial Registry (No. ChiCTR2100051191).

## RESULTS

### 1. Basic information of the training sets and the test sets

A total of 2,057 images (1,320 images of GISTs and 737 images of non-GISTs) from 367 patients (375 SELs, including 245 GISTs, 120 leiomyomas, 8 ectopic pancreas, 1 sclerotic fibroma, and 1 schwannoma) were finally chosen for analysis. The median age of the patients was 54 years

(range, 18 to 81 years), and 146 patients (39.78%) were male. A total of 1,851 images (1,188 images of GISTs and 663 images of non-GISTs) were randomly divided into the training sets and 206 images (132 images of GISTs and 74 images of non-GISTs) into the test sets. The characteristics of the SELs chosen are summarized in Table 1.

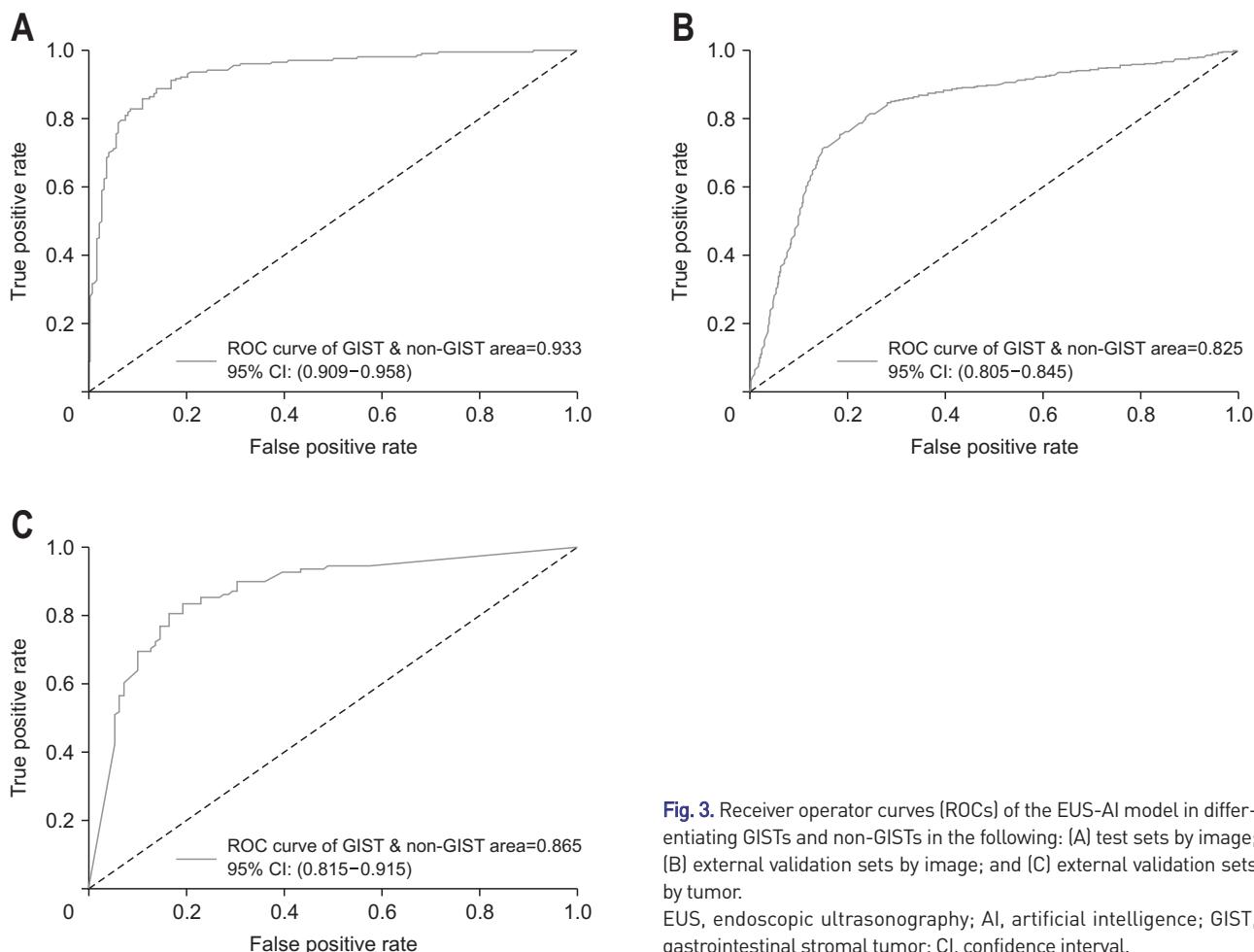
## 2. Performance of the EUS-AI model for prediction of GISTs and non-GISTs in the test sets

The sensitivity, specificity, PPV, NPV, and accuracy of the EUS-AI model in differentiating between GISTs and non-GISTs in the test sets by images were 84.85% (95% CI, 77.75% to 89.97%), 89.19% (95% CI, 80.09% to 94.42%), 93.33% (95% CI, 87.39% to 96.58%), 76.74% (95%

**Table 1.** Characteristics of the Subepithelial Lesions Chosen for Analysis

Characteristic	GISTs	Non-GISTs (n=130)
Location		
Cardia	2	5 (leiomyoma)
Gastric fundus	156	67 (leiomyoma)
Junction of gastric fundus and gastric body	2	1 (leiomyoma)
Gastric body	80	53 (47 leiomyomas, 5 ectopic pancreas, 1 sclerotic fibroma)
Gastric angle	2	1 (ectopic pancreas)
Gastric antrum	3	3 (2 ectopic pancreas, 1 schwannoma)
Total	245	130 (120 leiomyomas, 8 ectopic pancreas, 1 sclerotic fibroma, 1 schwannoma)
Size, No. (%)		
≥20 mm	74 (30.20)	28 (21.54)
<20 mm	171 (69.80)	102 (78.46)

GIST, gastrointestinal stromal tumor.



**Fig. 3.** Receiver operator curves (ROCs) of the EUS-AI model in differentiating GISTs and non-GISTs in the following: (A) test sets by image; (B) external validation sets by image; and (C) external validation sets by tumor.

EUS, endoscopic ultrasonography; AI, artificial intelligence; GIST, gastrointestinal stromal tumor; CI, confidence interval.

CI, 66.79% to 84.41%), and 86.41% (95% CI, 81.06% to 90.43%), respectively.

### 3. Basic information for the external validation sets

A total of 914 images (656 images of GISTs and 258 images of non-GISTs) from 106 patients (108 SELs, including 80 GISTs, 27 leiomyomas, and 1 schwannoma) were finally chosen in the external validation sets. The number of images selected from each hospital is presented in Supplementary Table 3.

### 4. Performance of the EUS-AI model for prediction of GISTs and non-GISTs in the external validation sets

The sensitivity, specificity, PPV, NPV, and accuracy of the EUS-AI model in differentiating GISTs and non-GISTs in the external validation sets by images were 82.01% (95% CI, 78.89% to 84.76%), 68.22% (95% CI, 62.30% to 73.60%), 86.77% (95% CI, 83.88% to 89.22%), 59.86% (95% CI, 54.17% to 65.31%), and 78.12% (95% CI, 75.32% to 80.68%), respectively. The sensitivity, specificity, PPV, NPV, and accuracy of the EUS-AI model for differentiating GISTs and non-GISTs in the external validation sets by tumors were 83.75% (95% CI, 74.16% to 90.25%), 71.43% (95% CI, 52.94% to 84.75%), 89.33% (95% CI, 80.34% to 94.50%), 60.61% (95% CI, 43.68% to 75.32%), and 80.56% (95% CI, 72.10% to 86.92%), respectively. The receiver operating characteristic curves for the EUS-AI model for discriminating GISTs and non-GISTs are shown in Fig. 3. The false positive or negative EUS images and the confusion matrices of the pairwise comparison in the external validation sets are presented in Fig. 4.

### 5. Subgroup analysis

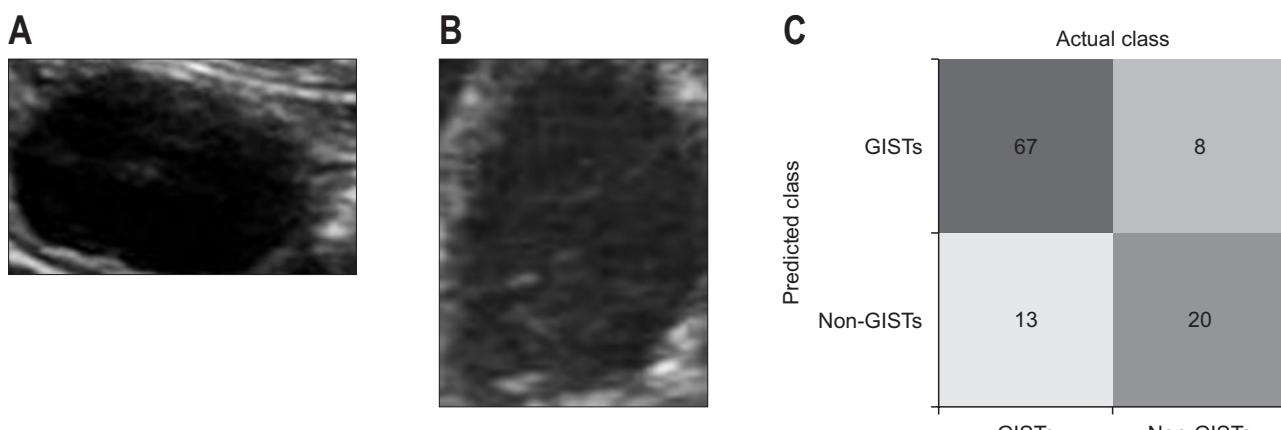
We further performed subgroup analysis using the test sets and external validation sets by lesion size and by hospital. The accuracy for SELs  $\geq 20$  mm in the external validation sets by images was 84.6%, and 87.5% by tumors; the accuracy for SELs  $< 20$  mm in the external validation sets by images was 69.33%, and 75% by tumors. The diagnostic performance of the EUS-AI model varied in different hospitals (Supplementary Table 4).

### 6. Performance of endosonographers for prediction of GISTs and non-GISTs with and without the aid of EUS-AI

The diagnostic performance of endosonographers, varied even among endosonographers with the same level of experience, especially with aspect to specificity (Supplementary Table 5 shows the results of the sets used to build the model, and Supplementary Table 6 shows the results of the external validation sets).

Both in the test sets and external validation sets, the EUS-AI model showed much better specificity than that of novices and some experts (except for expert 3, Table 2 shows the comparison result between the EUS-AI model and each endosonographer in the external validation sets by tumors).

With the aid of EUS-AI, the sensitivity of expert 3 and novice 1 showed a significant increase ( $p=0.043$  for expert 3,  $p=0.005$  for novice 1), while for other endosonographers, there were no significant difference (the comparison result is shown in Table 3). The specificity of expert 2 and novice 3 showed a significant increase ( $p=0.007$  for expert 2,  $p<0.001$  for novice 3), while no significant difference was observed for other endosonographers; the accuracy of expert 3, novice 1, and novice 3 increase ( $p=0.047$  for ex-



**Fig. 4.** (A) EUS image of a GIST misdiagnosed as non-GIST by EUS-AI. (B) EUS image of a non-GIST misdiagnosed as a GIST by EUS-AI. (C) Confusion matrices show the pairwise comparison (number of tumors) in the external validation sets.  
EUS, endoscopic ultrasonography; GIST, gastrointestinal stromal tumor; AI, artificial intelligence.

**Table 2.** Comparison of the Diagnostic Performance between the EUS-AI Model and Each Endosonographer in the External Validation Sets by Tumors

	Sensitivity, % (95% CI)	p-value	Specificity, % (95% CI)	p-value	Accuracy, % (95% CI)	p-value
AI	83.75 [74.16–90.25]	-	71.43 [52.94–84.75]	-	80.56 [72.10–86.92]	-
Expert 1	95.00 [87.84–98.04]	0.021	71.43 [52.94–84.75]	1.000	88.89 [81.58–93.53]	0.089
Expert 2	93.75 [86.19–97.30]	0.045	35.71 [20.71–54.17]	0.007	78.70 [70.07–85.37]	0.735
Expert 3	80.00 [69.95–87.30]	0.538	100 [87.94–100]	0.002	85.19 [77.28–90.67]	0.367
Novice 1	65.70 [56.64–76.76]	0.017	64.29 [45.83–79.29]	0.567	66.67 [57.34–74.85]	0.021
Novice 2	87.50 [78.50–93.07]	0.499	32.14 [17.93–50.67]	0.003	73.15 [64.10–80.61]	0.197
Novice 3	86.25 [77.03–92.15]	0.658	10.71 [3.71–27.20]	<0.001	66.67 [57.34–74.85]	0.021

EUS, endoscopic ultrasonography; AI, artificial intelligence; CI, confidence interval.

**Table 3.** Comparison of the Diagnostic Performance of the Endosonographers before and after the Use of EUS-AI in the External Validation Sets by Tumors

	Sensitivity, % (95% CI)	p-value	Specificity, % (95% CI)	p-value	Accuracy, % (95% CI)	p-value
Expert 1+AI	87.50 [78.50–93.07]	0.093	71.43 [52.94–84.75]	1.000	83.33 [75.19–89.19]	0.238
Expert 2+AI	85.00 [75.59–91.21]	0.073	71.43 [52.94–84.75]	0.007	81.48 [73.12–87.68]	0.609
Expert 3+AI	91.25 [83.02–95.70]	0.043	100 [87.94–100]	1.000	93.52 [87.22–96.83]	0.047
Novice 1+AI	86.25 [77.03–92.15]	0.005	64.29 [45.83–79.29]	1.000	80.56 [72.10–86.92]	0.021
Novice 2+AI	95.00 [87.84–98.04]	0.093	50.00 [32.63–67.37]	0.174	83.33 [75.19–89.19]	0.070
Novice 3+AI	88.75 [79.98–93.97]	0.633	57.14 [39.07–73.49]	<0.001	80.56 [72.10–86.92]	0.021

EUS, endoscopic ultrasonography; AI, artificial intelligence; CI, confidence interval.

pert 3, p=0.021 for novice 1 and novice 3), while there was no significant difference for other endosonographers.

## DISCUSSION

In this study, we built an EUS-AI model to predict the histopathological diagnosis of gastric SELs originating from muscularis propria based on EUS images. Due to the limited sample size of images for building the model in certain histopathology categories, we were only able to perform binary classification of GISTs and non-GISTs using this model. The model was built by images obtained from two centers, with various types of echoendoscopes and ultrasound systems, which are widely used in clinical centers. The performance of this EUS-AI model was validated in multicenter external validation. The sensitivity, specificity, and accuracy of the model in the external validation sets by tumors were 83.75%, 71.43%, and 80.56%, respectively. Compared with the endosonographers, this EUS-AI model showed superior specificity. This EUS-AI model was found to improve the diagnostic sensitivity, specificity, or accuracy of some endosonographers.

Previous studies have also tried to build EUS-AI models to differentiate SELs. The earliest published study was probably performed by Minoda *et al.*<sup>8</sup> Later, Kim *et al.*<sup>12</sup> also built a convolutional neural network-AI system to discriminate GISTs and non-GISTs, and its diagnostic perfor-

mance was similar to that of Minoda *et al.*<sup>8</sup> and our model. Other studies have tried to build EUS-AI models to differentiate GISTs and leiomyomas, but most of them were single-center studies, without external validation.<sup>8,9,12,13</sup>

Yang *et al.*<sup>10</sup> constructed an EUS-AI model to differentiate GISTs and leiomyomas with prospective and external validation. The sensitivity, specificity, and diagnostic accuracy of their model in the external validation sets were 45.8%, 84.6%, and 66.0%, respectively. The differences in the results may be attributed to some differences between the studies. First, Yang *et al.*<sup>10</sup> chose all GISTs and leiomyomas regardless of the location, while we only chose gastric SELs originating from muscularis propria, and we did not perform prospective validation. Second, in some studies, the labeled tumor images were resized to different sizes to reduce the influence of size in building the models.<sup>8,10,14</sup> In a previous study, the shapes of the SELs showed variability among lesions with different histopathological results;<sup>15</sup> therefore, we tried to trim the images into the shapes of the SELs, but the diagnostic performance was not that satisfactory. Third, the accuracy of the EUS-AI model may be influenced by the EUS probes, echoendoscopes, and EUS machines used. As we can see, in Fig. 4, the difference in some of the EUS images was small, and both the endosonographers and the EUS-AI model failed to make the correct classification. Fourth, in their study, only about 30% of the consecutive patients had a clear histopathological diagnosis, which caused verification bias, while in our

study, all the patients had confirmed histopathological diagnoses.

We performed two kinds of subgroup analysis, and the first one was by lesion size. We found that a tendency for higher sensitivity and accuracy of the EUS-AI for SELs no less than 20 mm compared to SELs <20 mm (sensitivity of the test sets by images, sensitivity and accuracy of the external validation by images had significant difference); however, there was no significant difference in specificity. The results were similar to those reported by Yang *et al.*<sup>10</sup> and Minoda *et al.*<sup>8</sup> The second subgroup analysis was performed by hospitals. There are several potential explanations for this phenomenon. First, though we used almost all kinds of echoendoscopes that were utilized in clinical settings to build the models, the proportion of each echoendoscope varied, and the echoendoscopes were different in the hospitals from the external validation sets, which may have influenced the result. Second, the diagnostic performance of the EUS-AI model was influenced by the quality of the images. Although we selected only good-quality images, we still found variable quality images in different hospitals.

In this study, the diagnostic performance varied among different endosonographers, even with the same level of experience, especially with respect to specificity. This indicates the need for introducing a system which may help reduce the inter-observer variability among the endosonographers. With the aid of the EUS-AI model, some endosonographers showed improvement in sensitivity, specificity, or accuracy. We further found that, for those with equal or higher sensitivity/specificity/accuracy, the EUS-AI model did not help much, but for those with low sensitivity/specificity/accuracy, the EUS-AI improved their performance. Hence, this EUS-AI model would be more helpful for those novices, as it might help improve their accuracy and reduce the inter-observer variability. On the other hand, whether this model can help the endosonographers not only depends on its own diagnostic performance, but also on the confidence of the endosonographer in the model. As for novice 2, his specificity was low, but he did not have much confidence in the EUS-AI model, so his specificity did not increase much.

In our EUS-AI model, we collected the normal B-mode images, and evaluation of the brightness or grayscale histogram, heterogeneity, shape, "halo" sign, and cystic changes in EUS images may help distinguish SELs.<sup>16-18</sup> However, there are other techniques that aid in the differentiation of SELs. A previous study demonstrated the usefulness of elastography in differentiating GISTs and non-GISTs, as

the former were found to be harder.<sup>19</sup> Contrast-enhanced EUS was also reported to be a complementary tool to differentiate GISTs from non-GISTs, as GISTs usually showed early, clear enhancement, and avascular areas in the center of the lesions (referring to necrosis), while non-GISTs showed little or no enhancement.<sup>20,21</sup> A recently published study used contrast-enhanced harmonic EUS images to build AI model to distinguish GISTs and leiomyomas, and the accuracy of AI model was comparable to that of experts.<sup>22</sup> The above new techniques have all been shown to be useful in distinguishing between GISTs and non-GISTs. Incorporation of the information or the images or videos obtained with these techniques for building a multi-modal EUS-AI model may yield much better results.

Some limitations of our study should be taken into consideration while interpreting the results. First, we only selected gastric SELs originating from muscularis propria to build the model, since this was our first step towards building an EUS-AI model for distinguishing the SELs, and the SELs are most frequently found in the stomach. Moreover, the situations of gastric SELs are more complicated.<sup>23</sup> Indeed, most of the gastric SELs originate from the muscularis propria,<sup>24</sup> so we chose this category to build the prototype of the model. Our subsequent work may include all SELs from the esophagus, duodenum, and large intestine, with all layers to build a modified model. Second, the number of images and tumors used to build the EUS-AI model can still be increased. AI is based upon big data, and inclusion of more images, especially other types of tumors rather than GISTs or leiomyomas, may enable multi-class classifications using AI. Third, for EUS examinations, various machines were used, and there might be differences in image resolution or quality among the machines, so we collected images of different lesions from varied hospitals and using different endoscopes to minimize the influence. Fourth, we did not perform prospective validation, and the validations were all based on images rather than videos. It would be better to validate this EUS-AI model with prospective videos prior to its application in clinical practice.

In conclusion, we developed an EUS-AI model to classify gastric SELs originating from muscularis propria into GISTs and non-GISTs with good accuracy. The model may help enhance the sensitivity, specificity, and accuracy of endosonographers in differentiating gastric SELs. Further work is required to develop a multi-modal EUS-AI system which incorporates the information of location, elastography, contrast-enhanced, and detective flow imaging techniques, to make more accurate and multi-class classifications.

## CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

## ACKNOWLEDGEMENTS

This study was funded by grants from the Sun Yat-sen University Clinical Research 5010 Program (grant number: 2014008), and the Sixth Affiliated Hospital of Sun Yat-sen University of Horizontal Program (grant number: H202101162024041054).

The authors thank all the members from Tianjin Economic-Technological Development Area (TEDA) Yujin Digestive Health Industry Research Institute who have made contributions to this program, and thank Chujun Li from the Sixth Affiliated Hospital, Sun Yat-sen University for his support.

## AUTHOR CONTRIBUTIONS

Study concept and design: J.S., M.Z. Data acquisition: Y. Lu, J.W., M.H., Q.Z., L.E., H.S., W.C., K.C., Y. Liu, B.Q., Q.X., J.S. Data analysis and interpretation: Y. Lu, J.W., M.H., W.C., G.L., J.S., Y.W., Y. Luo, J.M. Drafting of the manuscript: Y. Lu, J.W. Critical revision of the manuscript for important intellectual content: W.Z., J.S., M.Z. Statistical analysis: Y. Lu, J.W., M.H., Y.W., Y. Luo, J.M. Obtained funding: M.Z., C.L. Administrative, technical, or material support; study supervision: Y.W., Y. Luo, J.M., J.S., M.Z. Approval of final manuscript: all authors.

## ORCID

Yi Lu	<a href="https://orcid.org/0000-0003-3979-5421">https://orcid.org/0000-0003-3979-5421</a>
Jiachuan Wu	<a href="https://orcid.org/0000-0001-8490-2965">https://orcid.org/0000-0001-8490-2965</a>
Minhui Hu	<a href="https://orcid.org/0000-0001-9909-7356">https://orcid.org/0000-0001-9909-7356</a>
Qinghua Zhong	<a href="https://orcid.org/0000-0001-6199-5599">https://orcid.org/0000-0001-6199-5599</a>
Limian Er	<a href="https://orcid.org/0000-0001-7709-316X">https://orcid.org/0000-0001-7709-316X</a>
Huihui Shi	<a href="https://orcid.org/0000-0001-5500-6467">https://orcid.org/0000-0001-5500-6467</a>
Weihui Cheng	<a href="https://orcid.org/0000-0001-7411-3815">https://orcid.org/0000-0001-7411-3815</a>
Ke Chen	<a href="https://orcid.org/0000-0001-7680-381X">https://orcid.org/0000-0001-7680-381X</a>
Yuan Liu	<a href="https://orcid.org/0000-0002-3572-1380">https://orcid.org/0000-0002-3572-1380</a>
Bingfeng Qiu	<a href="https://orcid.org/0000-0001-8255-619X">https://orcid.org/0000-0001-8255-619X</a>
Qiancheng Xu	<a href="https://orcid.org/0000-0003-4435-9261">https://orcid.org/0000-0003-4435-9261</a>
Guangshun Lai	<a href="https://orcid.org/0000-0002-0855-0627">https://orcid.org/0000-0002-0855-0627</a>
Yufeng Wang	<a href="https://orcid.org/0000-0001-9080-7040">https://orcid.org/0000-0001-9080-7040</a>
Yuxuan Luo	<a href="https://orcid.org/0000-0002-4505-4972">https://orcid.org/0000-0002-4505-4972</a>

Jinbao Mu	<a href="https://orcid.org/0000-0002-9172-2804">https://orcid.org/0000-0002-9172-2804</a>
Wenjie Zhang	<a href="https://orcid.org/0000-0002-7757-9546">https://orcid.org/0000-0002-7757-9546</a>
Min Zhi	<a href="https://orcid.org/0000-0001-8178-5572">https://orcid.org/0000-0001-8178-5572</a>
Jiachen Sun	<a href="https://orcid.org/0000-0003-3646-1039">https://orcid.org/0000-0003-3646-1039</a>

## SUPPLEMENTARY MATERIALS

Supplementary materials can be accessed at <https://doi.org/10.5009/gnl220347>.

## REFERENCES

- McCarty TR, Ryou M. Endoscopic diagnosis and management of gastric subepithelial lesions. *Curr Opin Gastroenterol* 2020;36:530-537.
- Hedenbro JL, Ekelund M, Wetterberg P. Endoscopic diagnosis of submucosal gastric lesions: the results after routine endoscopy. *Surg Endosc* 1991;5:20-23.
- Dumonceau JM, Deprez PH, Jenssen C, et al. Indications, results, and clinical impact of endoscopic ultrasound (EUS)-guided sampling in gastroenterology: European Society of Gastrointestinal Endoscopy (ESGE) Clinical Guideline-updated January 2017. *Endoscopy* 2017;49:695-714.
- Faulx AL, Kothari S, Acosta RD, et al. The role of endoscopy in subepithelial lesions of the GI tract. *Gastrointest Endosc* 2017;85:1117-1132.
- Nishida T, Blay JY, Hirota S, Kitagawa Y, Kang YK. The standard diagnosis, treatment, and follow-up of gastrointestinal stromal tumors based on guidelines. *Gastric Cancer* 2016;19:3-14.
- Sakamoto H, Kitano M, Kudo M. Diagnosis of subepithelial tumors in the upper gastrointestinal tract by endoscopic ultrasonography. *World J Radiol* 2010;2:289-297.
- Landi B, Palazzo L. The role of endosonography in submucosal tumours. *Best Pract Res Clin Gastroenterol* 2009;23:679-701.
- Minoda Y, Ihara E, Komori K, et al. Efficacy of endoscopic ultrasound with artificial intelligence for the diagnosis of gastrointestinal stromal tumors. *J Gastroenterol* 2020;55:1119-1126.
- Oh CK, Kim T, Cho YK, et al. Convolutional neural network-based object detection model to identify gastrointestinal stromal tumors in endoscopic ultrasound images. *J Gastroenterol Hepatol* 2021;36:3387-3394.
- Yang X, Wang H, Dong Q, et al. An artificial intelligence system for distinguishing between gastrointestinal stromal tumors and leiomyomas using endoscopic ultrasonography. *Endoscopy* 2022;54:251-261.
- Zhang H, Wu C, Zhang Z, et al. ResNeSt: Split-Attention

Networks. arXiv:2004.08955 [Preprint]. [cited 2022 Jun 7]. Available from: <https://doi.org/10.48550/arXiv.2004.08955>

12. Kim YH, Kim GH, Kim KB, et al. Application of a convolutional neural network in the diagnosis of gastric mesenchymal tumors on endoscopic ultrasonography images. *J Clin Med* 2020;9:3162.
13. Seven G, Silahtaroglu G, Seven OO, Senturk H. Differentiating gastrointestinal stromal tumors from leiomyomas using a neural network trained on endoscopic ultrasonography images. *Dig Dis* 2022;40:427-435.
14. Hirai K, Kuwahara T, Furukawa K, et al. Artificial intelligence-based diagnosis of upper gastrointestinal subepithelial lesions on endoscopic ultrasonography images. *Gastric Cancer* 2022;25:382-391.
15. Seo SW, Hong SJ, Han JP, et al. Accuracy of a scoring system for the differential diagnosis of common gastric subepithelial tumors based on endoscopic ultrasonography. *J Dig Dis* 2013;14:647-653.
16. Khan S, Zhang R, Fang W, et al. Reliability of endoscopic ultrasound using miniprobes and grayscale histogram analysis in diagnosing upper gastrointestinal subepithelial lesions. *Gastroenterol Res Pract* 2020;2020:6591341.
17. Lee MW, Kim GH, Kim KB, et al. Digital image analysis-based scoring system for endoscopic ultrasonography is useful in predicting gastrointestinal stromal tumors. *Gastric Cancer*;22:980-987.
18. Vaicekauskas R, Urbonienė J, Stanaitis J, Valantinas J. Evaluation of upper endoscopic and endoscopic ultrasound features in the differential diagnosis of gastrointestinal stromal tumors and leiomyomas in the upper gastrointestinal tract. *Visc Med* 2020;36:318-324.
19. Tsuji Y, Kusano C, Gotoda T, et al. Diagnostic potential of endoscopic ultrasonography-elastography for gastric submucosal tumors: a pilot study. *Dig Endosc* 2016;28:173-178.
20. Lefort C, Gupta V, Lisotti A, et al. Diagnosis of gastric submucosal tumors and estimation of malignant risk of GIST by endoscopic ultrasound: comparison between B mode and contrast-harmonic mode. *Dig Liver Dis* 2022;53:1486-1491.
21. Pesenti C, Bories E, Caillol F, et al. Characterization of subepithelial lesions of the stomach and esophagus by contrast-enhanced EUS: a retrospective study. *Endosc Ultrasound* 2019;8:43-49.
22. Tanaka H, Kamata K, Ishihara R, et al. Value of artificial intelligence with novel tumor tracking technology in the diagnosis of gastric submucosal tumors by contrast-enhanced harmonic endoscopic ultrasonography. *J Gastroenterol Hepatol* 2022;37:841-846.
23. Deprez PH, Moons LMG, O'Toole D, et al. Endoscopic management of subepithelial lesions including neuroendocrine neoplasms: European Society of Gastrointestinal Endoscopy (ESGE) guideline. *Endoscopy* 2022;54:412-429.
24. Kim B, Kang S, Lee E, et al. Gastric subepithelial tumor: long-term natural history and risk factors for progression. *Surg Endosc* 2022;36:5232-5242.