

# Artificial intelligence in endoscopic ultrasonography: risk stratification of gastric gastrointestinal stromal tumors

Yi Lu<sup>\*</sup>, Lu Chen<sup>\*</sup>, Jiachuan Wu, Limian Er, Huihui Shi, Weihui Cheng, Ke Chen, Yuan Liu, Bingfeng Qiu, Qiancheng Xu, Yue Feng, Nan Tang, Fuchuan Wan, Jiachen Sun<sup>\*</sup> and Min Zhi

## Abstract

**Background:** Previous studies have identified useful endoscopic ultrasonography (EUS) features to predict the malignant potential of gastrointestinal stromal tumors (GISTs). However, the results of the studies were not consistent. Artificial intelligence (AI) has shown promising results in medicine.

**Objectives:** We aimed to build a risk stratification EUS-AI model to predict the malignancy potential of GISTs.

**Design:** This was a retrospective study with external validation.

**Methods:** We developed two models using EUS images from two hospitals to predict the GIST risk category. Model 1 was the four-category risk EUS-AI model, and Model 2 was the two-category risk EUS-AI model. The diagnostic performance of the models was validated with external cohorts.

**Results:** A total of 1320 images (880 were very low-risk, 269 were low-risk, 68 were intermediate-risk, and 103 were high-risk) were finally chosen for building the models and test sets, and a total of 656 images (211 were very low-risk, 266 were low-risk, 88 were intermediate-risk, and 91 were high-risk) were chosen for external validation. The overall accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) for the four-category risk EUS-AI model in the external validation sets by tumor were 74.50%, 55.00%, 79.05%, 53.49%, and 81.63%, respectively. The accuracy, sensitivity, specificity, PPV, and NPV for the two-category risk EUS-AI model for the prediction of very low-risk GISTs in the external validation sets by tumor were 86.25%, 94.44%, 79.55%, 79.07%, and 94.59%, respectively.

**Conclusion:** We developed a EUS-AI model for the risk stratification of GISTs with promising results, which may complement current clinical practice in the management of GISTs.

**Registration:** The study has been registered in the Chinese Clinical Trial Registry (No. ChiCTR2100051191).

**Keywords:** Artificial intelligence, endoscopic ultrasonography, gastric, gastrointestinal stromal tumors, risk

Received: 28 January 2023; revised manuscript accepted: 4 May 2023.

## Introduction

Subepithelial lesions (SELs) are most frequently found in the stomach, most of which are gastrointestinal stromal tumors (GISTs).<sup>1</sup> GISTs have

varying degrees of malignancy potential. For example, micro-GISTs (<10 mm) have almost no malignancy potential, whereas large GISTs and GISTs with a high mitotic count have a high

*Ther Adv Gastroenterol*

2023, Vol. 16: 1–10

DOI: 10.1177/  
17562848231177156

© The Author(s), 2023.  
Article reuse guidelines:  
sagepub.com/journals-  
permissions

Correspondence to:

**Jiachen Sun**  
Department of  
Gastrointestinal  
Endoscopy, Guangdong  
Provincial Key Laboratory  
of Colorectal and Pelvic  
Floor Diseases, The Sixth  
Affiliated Hospital, Sun  
Yat-sen University, 26  
Yuancun Erheng Road,  
Guangzhou 510655,  
People's Republic of China  
[sunjch8@mail.sysu.edu.cn](mailto:sunjch8@mail.sysu.edu.cn)

**Min Zhi**  
Department of  
Gastroenterology,  
Guangdong Provincial Key  
Laboratory of Colorectal  
and Pelvic Floor Diseases,  
The Sixth Affiliated  
Hospital, Sun Yat-sen  
University, 26 Yuancun  
Erheng Road, Guangzhou  
510655, People's Republic  
of China  
[zhimin@mail.sysu.edu.cn](mailto:zhimin@mail.sysu.edu.cn)

**Yi Lu**  
Department of  
Gastrointestinal  
Endoscopy, Guangdong  
Provincial Key Laboratory  
of Colorectal and Pelvic  
Floor Diseases, The  
Sixth Affiliated Hospital,  
Sun Yat-sen University,  
Guangzhou, People's  
Republic of China

**Lu Chen**  
Department of Internal  
Medicine, Advent Health  
Palm Coast, Palm Coast,  
FL, USA

**Jiachuan Wu**  
Digestive Endoscopy  
Center, Guangdong  
Second Provincial General  
Hospital, Guangzhou,  
People's Republic of China

**Limian Er**  
**Huihui Shi**  
Department of Endoscopy,  
The Fourth Hospital of  
Hebei Medical University,  
Shijiazhuang, People's  
Republic of China

**Weihui Cheng**  
Department of  
Gastroenterology,  
Yangjiang Hospital of  
Traditional Chinese  
Medicine, Yangjiang,  
People's Republic of China

**Ke Chen**  
**Yuan Liu**  
Department of Endoscopy,  
Fudan University Shanghai  
Cancer Center, Shanghai,  
People's Republic of China

**Bingfeng Qiu**  
**Qiancheng Xu**  
Department of  
Gastroenterology,  
Zhoushan Hospital  
of Zhejiang Province,  
Zhoushan, People's  
Republic of China

**Yue Feng**  
Tianjin Economic-  
Technological  
Development Area  
(TEDA) Yujin Digestive  
Health Industry Research  
Institute, Tianjin, People's  
Republic of China

**Nan Tang**  
Tianjin Center for Medical  
Devices Evaluation and  
Inspection, Tianjin,  
People's Republic of China

**Fuchuan Wan**  
Tianjin Economic-  
Technological  
Development Area (TEDA)  
Yujin Artificial Intelligence  
Medical Technology Co,  
Ltd, Tianjin, People's  
Republic of China

\*These authors  
contributed equally.

metastatic probability and recurrence rate.<sup>2</sup> Numerous systems have been proposed for the risk stratification of GISTs. The most widely used systems are the National Institute of Health (NIH) criteria, modified NIH criteria, and Armed Forces Institute of Pathology (AFIP) criteria.<sup>3-6</sup> The variables that are included in the abovementioned systems are mainly tumor size, mitotic index, and tumor location.

The mitotic index is obtained via histopathologic sections of the resected tumor or endoscopic ultrasonography-guided fine needle aspiration or biopsy (EUS-FNA/B). If we are able to predict the malignancy potential of a suspected GIST, it could aid clinical decision-making, especially for small GISTs. Previous studies have attempted to identify useful EUS features to predict the malignant potential of GISTs, and they found that, large size, cystic change, surface ulceration, extraluminal border, depth, heterogeneity irregular borders, and a nonoval shape were potential predictors for malignancy.<sup>7-11</sup> However, these predictors were not consistent in the different studies; moreover, except for tumor size, the judgment of other predictors was somewhat subjective. Artificial intelligence (AI) via deep convolutional neural networks (DCNNs) has shown promising results in medicine.<sup>12-14</sup> Seven *et al.*<sup>15</sup> built a deep learning algorithm using EUS images to predict the malignancy potential of gastric GISTs with good accuracy. Their studies used the AFIP criteria as the reference, and the sample size was relatively small and lacked external validation. A previous study indicated that the modified NIH criteria were the best criteria to identify a single high-risk group for consideration of adjuvant therapy.<sup>2</sup> Hence, we aimed to build a risk stratification EUS-AI model to predict the malignancy potential of GISTs in reference to the modified NIH criteria with larger sample sizes and external validation.

## Methods

### *Clinical data and EUS image collection*

The EUS images used to build the risk stratification EUS-AI model and used for external validation were from our previous study, in which we successfully built a EUS-AI model to differentiate GISTs and non-GISTs with good accuracy.<sup>16</sup> EUS images obtained from The Sixth Affiliated Hospital, Sun Yat-sen University, and Guangdong

Second Provincial Central Hospital were consecutively collected to build the model. For external validation, we collected 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). The inclusion criteria for the collected images were as follows: good quality EUS images showing the tumors; confirmed histopathology of gastric GISTs (obtained by endoscopic resection, surgery, or FNA); and modified NIH criteria risk category could be obtained. The exclusion criteria included the following: poor-quality images and repeated images. The following information about the patients was also collected: age, sex, EUS results, and histopathology results. The collected images were categorized based on the modified NIH criteria risk category.

### *EUS procedures*

EUS was performed by experienced endosonographers with experience in more than 500 cases of SELs. The echoendoscopes that were used in the models and the external validation sets are shown in Supplemental Tables S1 and S2. The utilized frequency was 5–20 MHz.

### *Image processing and development of the deep learning models*

The image processing, augmentation, and development of the deep learning models were supported by the Tianjin Jinyu Artificial Intelligence Medical Technology Co., Ltd. The image processing was the same as in our previous study. After selecting the qualified images, two experts in EUS marked the borderline of the tumor with LabelMe software (a polygonal and open annotation tool developed by the Massachusetts Institute of Technology, Computer Science and Artificial Intelligence Laboratory), and the marked tumors were regarded as the regions of interest (ROIs). Afterward, the engineers trimmed the images to squares or rectangles by precisely fitting the ROI. For images with measuring lines or measuring marks, Adobe Photoshop (Version 13.0; Adobe Systems Software Ireland Ltd., Dublin, Ireland) was used to erase the lines or marks and to preserve the originality of the images as much as possible. Image augmentation technology, such as mirror flip, horizontal flip, and rotation in certain

degrees of the EUS images, was applied. The pre-processed images were then changed into the RGB three-channel to be generated as the model input.

We developed two models in this study. Model 1 was a four-category risk EUS-AI model, and Model 2 was a two-category risk EUS-AI model. In Model 1, we classified the images into four categories based on the modified NIH criteria: very low-, low-, intermediate-, and high-risk. In Model 2, we classified the images into two categories: very low- and non-very low-risk (the latter included low-, intermediate-, and high-risk categories). The process was mainly implemented by Python (Version 3.7; Python Software Foundation, Wilmington, DE, USA) and PyTorch (Version 1.7.1; Facebook artificial intelligence research institute, Menlo Park, CA, USA). The chosen images were randomly divided into training sets and test sets at a ratio of 7.5:2.5, and 10-fold cross-validation was used.

We used the DCNN classifier known as ResNeSt-50 to train Model 1 (Amazon and University of California, Davis, Davis, CA, USA). ResNeSt was developed by Zhang *et al.*<sup>17</sup>, and it showed superiority in accuracy and latency trade-off on image classification when compared with other backbones.<sup>17</sup> The Adam optimizer was used, which is a popular optimizer that combines the ideas of AdaGrad and RMSProp.<sup>18</sup> The initial learning rate was set as 0.003, and cosine annealing was adopted as the attenuation method. This method took the cosine function as a period and reset the learning rate at the maximum value of each period.<sup>19</sup> The model was trained for 300 epochs, and the final learning rate was  $3e-7$ . When the training reached 300 epochs, the training was stopped.

As Model 1 was a four-category model (which was a multicategory task), a deeper learning model was needed to learn the characteristics. Model 2 was a two-category model, and the use of ResNeSt-34 was sufficient to train the model. The Adam optimizer was subsequently used. The initial learning rate was set as 0.003, and cosine annealing was adopted as the attenuation method. The model was trained for 500 epochs, and the final learning rate was  $9e-8$ . When the training reached 500 epochs, the training was stopped.

The output of the model was the risk category (ranging from 0 to 100%) of the GISTs. The type with the higher probability was interpreted as the final risk category.

### *Statistical analysis*

Data are presented as the mean (standard deviation, SD) for normally distributed continuous variables and presented as the median (range) for nonnormally distributed continuous variables. Categorical variables are expressed as numbers (percentages). A receiver operating characteristic (ROC) curve was plotted, and the area under the curve (AUC) was calculated. The sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy of the risk category of the EUS-AI model, as well as the respective 95% confidence intervals (CIs), were calculated. The abovementioned calculation was performed by using the Scikit-learn package in Python. The reporting of this study conforms to the Standards for Reporting Diagnostic Accuracy (STARD) 2015 statement.<sup>20</sup> The study has been registered in the Chinese Clinical Trial Registry (No. ChiCTR2100051191).

## **Results**

### *Basic information about the training sets and the test sets*

A total of 1320 images (the number of images for very low-risk, low-risk, intermediate-risk, and high-risk was 880, 269, 68, and 103, respectively) from 243 patients (245 GIST cases, of which 168 were very low-risk, 56 were low-risk, 12 were intermediate-risk, and 9 were high-risk) were chosen for analysis. The median age of the patients was 56 (29–81) years, and 104 patients (42.80%) were male. A total of 990 images were randomly divided into the training sets, and 330 images were randomly divided into the test sets. The characteristics of the chosen GISTs are summarized in Table 1.

### *Basic information for the external validation sets*

A total of 656 images (the number of images for very low-risk, low-risk, intermediate-risk, and high-risk was 211, 266, 88, and 91, respectively) from 79 patients (80 GISTs, of which 36 were

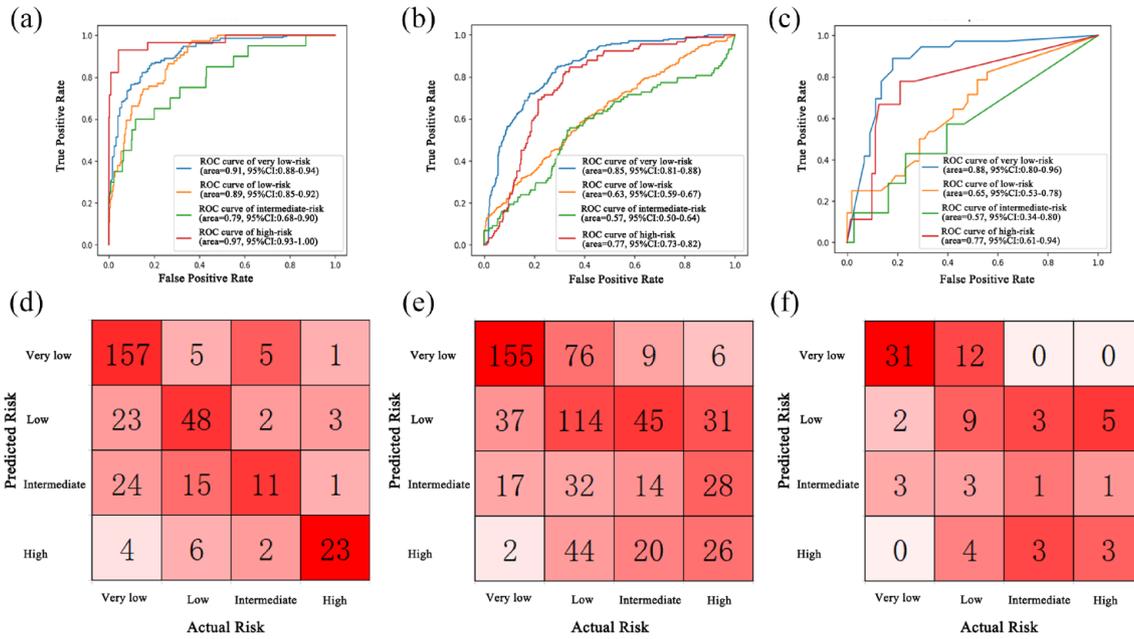
**Table 1.** The characteristics of the GISTs chosen for analysis.

Category	Very low	Low	Intermediate	High	Total
Training and test set					
Total	168	56	12	9	245
Location (n, %)					
Cardia	2 (1.19)	0	0	0	2 (0.82)
Gastric fundus	116 (69.05)	28 (50)	7 (58.33)	5 (55.56)	156 (63.67)
Junction of gastric fundus and gastric body	1 (0.6)	0	1 (8.33)	0	2 (0.82)
Gastric body	49 (29.17)	23 (41.07)	4 (33.33)	4 (44.44)	80 (32.65)
Gastric angle	0	2 (3.57)	0	0	2 (0.82)
Gastric antrum	0	3 (5.36)	0	0	3 (1.22)
Size (n, %)					
≥20 mm	0	54 (96.43)	11 (91.67)	9 (100)	74 (30.20)
<20 mm	168 (100)	2 (3.57)	1 (8.33)	0	171 (69.80)
External validation set					
Total	36	28	7	9	80
Location (n, %)					
Cardia	2 (5.56)	1 (3.57)	0	1 (11.11)	4 (5)
Gastric fundus	18 (50)	12 (42.86)	2 (28.57)	3 (33.33)	35 (43.75)
Junction of gastric fundus and gastric body	1 (2.78)	0	1 (14.29)	1 (11.11)	3 (3.75)
Gastric body	14 (38.89)	13 (46.43)	4 (57.14)	4 (44.44)	35 (43.75)
Gastric antrum	1 (2.78)	2 (7.14)	0	0	3 (3.75)
Size (n, %)					
≥20 mm	0	26 (92.86)	6 (85.71)	8 (88.89)	40 (50)
<20 mm	36 (100)	2 (7.14)	1 (14.29)	1 (11.11)	40 (50)
GIST, gastrointestinal stromal tumor.					

very low-risk, 28 were low-risk, 7 were intermediate-risk, and 9 were high-risk) were finally chosen in the external validation sets (Table 1). The number of images and tumors selected from each hospital are presented in Supplemental Table S3.

*Diagnostic performance of the four-category risk EUS-AI model for prediction of the risk category in the test sets*

The accuracy, sensitivity, specificity, PPV, and NPV for the four-category risk EUS-AI model in the test sets by images were 81.21%, 75.48%,



**Figure 1.** ROCs of the four-category risk EUS-AI model for prediction of the GISTs in the: (a) test sets by images, (b) external validation sets by images, (c) external validation sets by tumors, and confusion matrices of the pairwise comparison in the (d) test sets by images, (e) external validation sets by images, and (f) external validation sets by tumors.

AI, artificial intelligence; EUS, endoscopic ultrasonography; GIST, gastrointestinal stromal tumor; ROCs, receiver operating characteristic.

90.98%, 93.45%, and 68.52%, respectively, for very low-risk; 83.64%, 64.86%, 89.06%, 63.16%, and 89.76%, respectively, for low-risk; 85.15%, 55%, 87.1%, 21.57%, and 96.77%, respectively, for intermediate-risk; and 94.85%, 82.14%, 96.03%, 65.71%, and 98.31%, respectively, for high-risk (Supplemental Table S4). The overall accuracy, sensitivity, specificity, PPV, and NPV were 83.15%, 72.42%, 90.75%, 79.95%, and 77.52%, respectively. The ROC curves for each type versus other types are shown in Figure 1(a), and the confusion matrices showing the pairwise comparison (number of images) in the test sets are presented in Figure 1(d).

*Diagnostic performance of the four-category risk EUS-AI model for prediction of the risk category in the external validation sets*

The accuracy, sensitivity, specificity, PPV, and NPV for the four-category risk EUS-AI model in the external validation sets by images and by tumors are shown in Table 2. The overall accuracy, sensitivity, specificity, PPV, and NPV by images were 70.55%, 47.10%, 78.23%, 46.61%, and 77.88%, respectively. The overall accuracy, sensitivity, specificity, PPV, and NPV by tumor

were 74.50%, 55.00%, 79.05%, 53.49%, and 81.63%, respectively. The ROC curves and the confusion matrices showing the pairwise comparison results are presented in Figure 1.

*Diagnostic performance of the two-category risk EUS-AI model for prediction of the risk category in the test sets*

In Model 2, we classified the images into two categories: very low and non-very low-risk. The accuracy, sensitivity, specificity, PPV, and NPV for the two-category risk EUS-AI model for the prediction of very low-risk GISTs in the test sets by images were 90.91%, 90.38%, 91.80%, 94.95%, and 84.85%, respectively (Table 3). The ROC curve is presented in Figure 2(a).

*Diagnostic performance of the two-category risk EUS-AI model for prediction of the risk category in the external validation sets*

The accuracy, sensitivity, specificity, PPV, and NPV for the two-category risk EUS-AI model for the prediction of very low-risk GISTs in the external validation sets by image were 82.77%, 85.31%,

**Table 2.** Diagnostic performance of the four-category risk EUS-AI model for prediction of the risk category of GISTs in the external validation sets.

Risk stratification	Very low	Low	Intermediate	High
By images				
Accuracy	77.59% (74.25–80.62%)	59.6% (55.80–63.29%)	76.98% (73.61–80.04%)	80.03% (76.80–82.91%)
Sensitivity	73.46% (67.12–78.96%)	42.86% (37.05–48.86%)	15.91% (9.72–24.95%)	28.57% (20.31–38.57%)
Specificity	79.55% (75.56–83.04%)	71.03% (66.34–75.31%)	86.44% (83.38–89.02%)	88.32% (85.41–90.71%)
PPV	63.01% (56.82–68.80%)	50.22% (43.77–56.67%)	15.38% (9.39–24.18%)	28.26% (20.08–38.19%)
NPV	86.34% (82.68–89.33%)	64.57% (59.93–68.95%)	86.9% (83.87–89.44%)	88.48% (85.58–90.85%)
By tumors				
Accuracy	78.75% (68.58–86.29%)	63.75% (52.81–73.43%)	83.75% (74.16–90.25%)	83.75% (74.16–90.25%)
Sensitivity	86.11% (71.34–93.92%)	32.14% (17.93–50.66%)	14.29% (2.57–51.31%)	33.33% (12.06–64.58%)
Specificity	72.73% (58.15–83.65%)	80.77% (68.10–88.29%)	90.41% (81.50–95.28%)	90.14% (81.02–95.14%)
PPV	72.09% (57.31–82.35%)	47.37% (27.33–68.29%)	12.5% (2.24–47.09%)	30% (10.78–60.32%)
NPV	86.49% (72.02–94.09%)	68.85% (56.41–79.06%)	91.67% (82.99–96.12%)	91.43% (82.53–96.01%)
AI, artificial intelligence; EUS, endoscopic ultrasonography; GIST, gastrointestinal stromal tumor; NPV, negative predictive value; PPV, positive predictive value.				

**Table 3.** Diagnostic performance of the two-category risk EUS-AI model for prediction of the very low-risk GISTs in different sets.

Risk stratification	Test sets by images	External validation sets by images	External validation sets by tumors
Accuracy	90.91% (87.32–93.56%)	82.77% (79.70–85.47%)	86.25% (77.03–92.15%)
Sensitivity	90.38% (85.62–93.69%)	85.31% (79.90–89.45%)	94.44% (81.86–98.46%)
Specificity	91.80% (85.57–95.49%)	81.57% (77.71–84.90%)	79.55% (65.50–88.85%)
PPV	94.95% (90.95–97.23%)	68.70% (62.85–74.01%)	79.07% (64.79–88.58%)
NPV	84.85% (77.75–89.97%)	92.13% (89.05–94.40%)	94.59% (82.30–98.50%)
AI, artificial intelligence; EUS, endoscopic ultrasonography; GIST, gastrointestinal stromal tumor; NPV, negative predictive value; PPV, positive predictive value.			

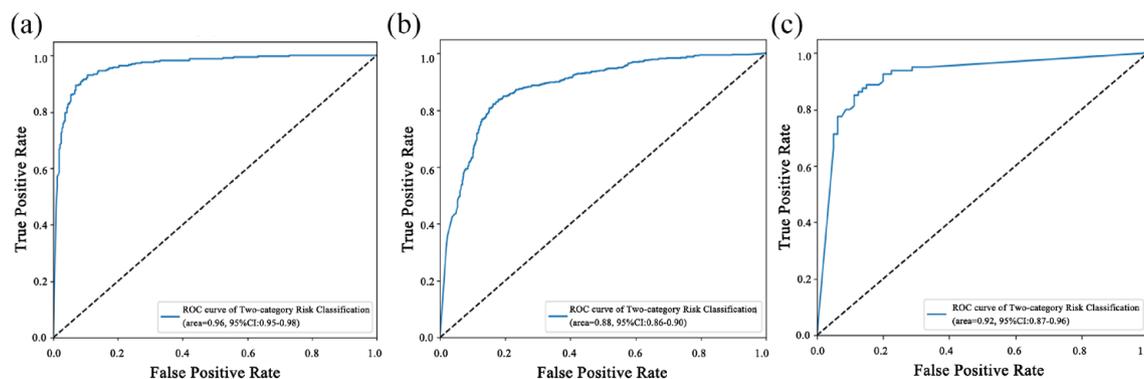
81.57%, 68.70%, and 92.13%, respectively, and those by tumor were 86.25%, 94.44%, 79.55%, 79.07%, and 94.59%, respectively (Table 3). The ROC curves are presented in Figure 2(b) and (c).

### Discussion

In this study, we built two risk stratification EUS-AI models to predict the malignant

potential of GISTs in reference to the modified NIH criteria, and the accuracy of the four-category risk EUS-AI model was not significantly good; however, the accuracy of the two-category risk EUS-AI model was much better with external validation.

Risk stratification of GISTs tries to evaluate the risk of poor outcomes and to choose patients who



**Figure 2.** ROCs of the two-category risk EUS-AI model for prediction of the very low-risk GISTs in the: (a) test sets by images, (b) external validation sets by images, and (c) external validation sets by tumors. AI, artificial intelligence; EUS, endoscopic ultrasonography; GIST, gastrointestinal stromal tumor; ROCs, receiver operating characteristic.

may benefit from adjuvant therapy.<sup>21</sup> Independent prognostic factors for GIST include the mitotic index, tumor size, tumor location (gastric *versus* non-gastric), and tumor rupture.<sup>2</sup> Patients in the low-risk group generally have favorable outcomes and may not require adjuvant therapy or frequent surveillance. In the past few years, several studies have explored AI applications in GISTs which not only include differentiating GISTs from non-GISTs,<sup>13,14,22–24</sup> but also risk stratification and prediction of prognosis.<sup>15</sup> Most of the previous studies regarding AI applications in GIST risk stratification were computed tomography (CT)-derived analyses and only a few studies have researched the performance of EUS-derived images. Two major machine learning methodologies, radiomic analysis and DCNNs have shown promise in pattern classification in previous studies.<sup>15,25</sup>

Seven *et al.*<sup>15</sup> investigated AI via DCNNs in predicting the malignant potential of GISTs. The overall sensitivity, specificity, and accuracy of the AI system for predicting four-category malignancy risk based on AFIP criteria in their study were 83%, 94%, and 82%, respectively, in the training dataset, and 75%, 73%, and 66%, respectively, in the validation cohort. When patients were divided into two categories: low-risk (including very low-risk and low-risk groups) and high-risk groups (including intermediate-risk and high-risk groups), the sensitivity, specificity, and accuracy in the validation cohort increased to 99.7%, 99.7%, and 99.6%, respectively.

In our study, the overall accuracy, sensitivity, and specificity of the four-category risk EUS-AI model

by tumor in the external validation set were 74.50%, 55.00%, and 79.05%, respectively, which is comparable to the results of the validation cohort in Seven's study. The overall performance of AI in the four-category risk EUS-AI model was not satisfactory because the sensitivity was observed to be relatively low. One possible reason for this phenomenon was that, in our study, the distribution of patients in each risk group of the training set was not balanced, and there were more patients in the very low-risk group than in the other three groups. Therefore, we may not have obtained optimized results for unbalanced classes, as the model did not obtain a sufficient understanding of the underlying classes. This imbalance is naturally inherent in our real-life medical practice. Intermediate and high-risk GISTs were usually larger in size and accompanied by ulceration which could be diagnosed by biopsies alone in most cases; thus, EUS was not required in the diagnosis and preoperative management. Additionally, some of the patients with intermediate- or high-risk GISTs underwent surgery directly after CT examinations with no additional required EUS in our center. These factors can explain the low number of patients in the higher-risk group.

When patients were divided into very low-risk and non-very low-risk groups according to modified NIH criteria, the two-category EUS-AI model that we built showed better accuracy, sensitivity, and specificity of 86.25%, 94.44%, and 79.55%, respectively, by tumor in the external validation sets. Previous studies<sup>15,25</sup> have combined the very low-risk and low-risk groups

together, which was different from our study. According to European guidelines,<sup>26</sup> very low-risk GISTs probably do not require routine follow-up. National Comprehensive Cancer Network guideline 2022 version<sup>27</sup> also mentioned that less frequent surveillance is acceptable for very small tumors (<2 cm), which is consistent with the modified NIH criteria for the very low-risk group. These results indicate that the very low-risk group has a much lower risk of recurrence compared to the other risk groups; therefore, dividing patients into very low-risk and non-very low-risk groups may be of more clinical significance because the management and follow-up strategy would be different. Our EUS-AI models demonstrated encouraging results in differentiating very low-risk GISTs from other risk groups, which can benefit actual clinical work.

In the study of Li *et al.*,<sup>25</sup> an EUS-derived radiomics model was developed to differentiate GISTs of the higher-risk classification (intermediate-risk and high-risk) from the lower-risk classification (very low-risk and low-risk). The accuracy, sensitivity, and specificity were 82.3%, 81.3%, and 82.6%, respectively. However, they did not specify which classification criteria were used in their study. The radiomics model has been investigated in many previous studies for the risk stratification of GISTs; however, most of the studies<sup>28</sup> were focused on the performance of AI using CT images. Li's study demonstrated that a EUS-derived radiomics model could increase preoperative diagnostic accuracy and provide a valuable reference in physicians' decision-making. The steps of radiomics analysis include the first extraction of numerous handcrafted imaging features, followed by feature selection and machine learning-based classification. However, handcrafted features are limited to the current knowledge of medical imaging, which may limit the potential of the predictive model.

Deep learning improves these handcrafted features by automatically learning discriminative features directly from images.<sup>29</sup> A previous study<sup>30</sup> demonstrated that a hybrid structure that includes different features selected with a radiomics model and DCNNs (AUC: 0.882, 95% CI: 0.816–0.947) outperforms independent radiomics (AUC: 0.807, 95% CI: 0.724–0.892) and CNNs (AUC: 0.826, 95% CI: 0.795–0.856) approaches for the classification of GISTs at CT. Moreover, deep learning also has

advantages and applications in other related fields, such as sequence methods and synthesis generation. Future studies that explore AI models integrating radiomics and DCNN features for the risk classification of GISTs using EUS images would likely add valuable information to this topic.

There were several limitations of our study. First, this was a retrospective study, and we did not perform prospective validation. It would be better to validate the EUS-AI model in a prospective cohort prior to its application in clinical practice. Another shortcoming, as was mentioned above, is that the number of intermediate- and high-risk GISTs was too small relative to very low-risk GISTs. In future studies, we need to enroll more patients with non-very low-risk to improve the imbalanced dataset model and to make it more stable. Despite these limitations, our results are of value in that our EUS-AI model was trained on a relatively large dataset and validated in an external validation cohort including four hospitals.

In conclusion, we developed a EUS-AI model for the risk stratification of GISTs with promising results. The accuracy and sensitivity of the two-category risk EUS-AI model were high with external validation, which may complement current clinical practice in the management of GISTs.

## Declarations

### *Ethics approval and consent to participate*

This study was a part of our previous work, which was approved by the Institutional Review Board of the Sixth Affiliated Hospital, Sun Yat-sen University (Approval Code: 2021ZSLYEC-319), and the informed consent was waived.

### *Consent for publication*

Not applicable.

### *Author contributions*

**Yi Lu:** Data curation; Formal analysis; Investigation; Project administration; Writing – original draft.

**Lu Chen:** Data curation; Formal analysis; Writing – original draft.

**Jiachuan Wu:** Data curation; Formal analysis; Investigation; Project administration.

**Limian Er:** Data curation; Investigation; Project administration; Supervision.

**Huihui Shi:** Data curation; Investigation; Project administration.

**Weihui Cheng:** Data curation; Investigation; Project administration.

**Ke Chen:** Data curation; Investigation; Project administration; Supervision.

**Yuan Liu:** Data curation; Investigation; Project administration.

**Bingfeng Qiu:** Data curation; Investigation; Project administration; Supervision.

**Qiancheng Xu:** Data curation; Investigation; Project administration.

**Yue Feng:** Formal analysis; Methodology; Project administration; Software; Validation; Visualization.

**Nan Tang:** Formal analysis; Methodology; Project administration; Software; Validation; Visualization.

**Fuchuan Wan:** Formal analysis; Methodology; Software; Validation; Visualization.

**Jiachen Sun:** Conceptualization; Data curation; Project administration; Resources; Writing – review & editing.

**Min Zhi:** Conceptualization; Funding acquisition; Supervision; Writing – review & editing.

### Acknowledgements

The authors thank all the members of Tianjin Economic-Technological Development Area (TEDA) Yujin Digestive Health Industry Research Institute who have made contributions to this program.

### Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: 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 No. H202101162024041054).

### Competing interests

The authors declare that there is no conflict of interest.

### Availability of data and materials

Data used in this study are available from the corresponding authors upon reasonable request.

### ORCID iDs

Yi Lu  <https://orcid.org/0000-0003-3979-5421>

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

### Supplemental material

Supplemental material for this article is available online.

### References

1. 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.
2. Joensuu H, Vehtari A, Riihimäki J, *et al.* Risk of recurrence of gastrointestinal stromal tumour after surgery: an analysis of pooled population-based cohorts. *Lancet Oncol* 2012; 13: 265–274.
3. Joensuu H. Risk stratification of patients diagnosed with gastrointestinal stromal tumor. *Hum Pathol* 2008; 39: 1411–1419.
4. Khoo CY, Chai X, Quek R, *et al.* Systematic review of current prognostication systems for primary gastrointestinal stromal tumors. *Eur J Surg Oncol* 2018; 44: 388–394.
5. Miettinen M and Lasota J. Gastrointestinal stromal tumors: pathology and prognosis at different sites. *Semin Diagn Pathol* 2006; 23: 70–83.
6. Fletcher CDM, Berman JJ, Corless C, *et al.* Diagnosis of gastrointestinal stromal tumors: a consensus approach. *Hum Pathol* 2002; 33: 459–465.
7. Jeon SW, Park YD, Chung YJ, *et al.* Gastrointestinal stromal tumors of the stomach: endosonographic differentiation in relation to histological risk. *J Gastroenterol Hepatol* 2007; 22: 2069–2075.
8. Shah P, Gao F, Edmundowicz SA, *et al.* Predicting malignant potential of gastrointestinal stromal tumors using endoscopic ultrasound. *Dig Dis Sci* 2009; 54: 1265–1269.

9. Chen TH, Hsu CM, Chu YY, *et al.* Association of endoscopic ultrasonographic parameters and gastrointestinal stromal tumors (GISTs): can endoscopic ultrasonography be used to screen gastric GISTs for potential malignancy? *Scand J Gastroenterol* 2016; 51: 374–377.
10. Chen T, Xu L, Dong X, *et al.* The roles of CT and EUS in the preoperative evaluation of gastric gastrointestinal stromal tumors larger than 2 cm. *Eur Radiol* 2019; 29: 2481–2489.
11. Brand B, Oesterhelweg L, Binmoeller KF, *et al.* Impact of endoscopic ultrasound for evaluation of submucosal lesions in gastrointestinal tract. *Dig Liver Dis* 2002; 34: 290–297.
12. Marya NB, Powers PD, Chari ST, *et al.* Utilisation of artificial intelligence for the development of an EUS-convolutional neural network model trained to enhance the diagnosis of autoimmune pancreatitis. *Gut* 2021; 70: 1335–1344.
13. Seven G, Silahtaroglu G, Seven OO, *et al.* Differentiating gastrointestinal stromal tumors from leiomyomas using a neural network trained on endoscopic ultrasonography images. *Dig Dis* 2022; 40: 427–435.
14. 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.
15. Seven G, Silahtaroglu G, Kochan K, *et al.* Use of artificial intelligence in the prediction of malignant potential of gastric gastrointestinal stromal tumors. *Dig Dis Sci* 2022; 67: 273–281.
16. Lu Y, Wu J, Hu M, *et al.* Artificial intelligence in the prediction of gastrointestinal stromal tumors on endoscopic ultrasonography images: development, validation and comparison with endosonographers. *Gut Liver* Epub ahead of print 26 January 2023. DOI: 10.5009/gnl220347.
17. Zhang H, Wu C, Zhang Z, *et al.* ResNeSt: split-attention networks, <https://arxiv.org/pdf/2004.08955.pdf>, 2020.
18. Cortiñas-Lorenzo B and Pérez-González F. Adam and the ants: on the influence of the optimization algorithm on the detectability of DNN watermarks. *Entropy* 2020; 22: 1–36.
19. Xie H and Wu Z. A robust fabric defect detection method based on improved RefineDet. *Sensors* 2020; 20: 1–24.
20. Korevaar DA, Cohen JF, Reitsma JB, *et al.* Updating standards for reporting diagnostic accuracy: the development of STARD 2015. *Res Integr Peer Rev* 2016; 1: 7.
21. Tirumani SH, Baheti AD, Tirumani H, *et al.* Update on gastrointestinal stromal tumors for radiologists. *Korean J Radiol* 2017; 18: 84–93.
22. 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: 1–11.
23. 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.
24. 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.
25. Li X, Jiang F, Guo Y, *et al.* Computer-aided diagnosis of gastrointestinal stromal tumors: a radiomics method on endoscopic ultrasound image. *Int J Comput Assist Radiol Surg* 2019; 14: 1635–1645.
26. Casali PG and Blay JY. Gastrointestinal stromal tumours: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2022; 33: 20–33.
27. National Comprehensive Cancer Network. Gastrointestinal stromal tumors (GISTs) (Version 1.2022), [https://www.nccn.org/professionals/physician\\_gls/pdf/gist.pdf](https://www.nccn.org/professionals/physician_gls/pdf/gist.pdf) (accessed 21 January 2022)
28. Yang CW, Liu XJ, Liu SY, *et al.* Current and potential applications of artificial intelligence in gastrointestinal stromal tumor imaging. *Contrast Media Mol Imaging* 2020; 2020: 6058159.
29. Sun Q, Lin X, Zhao Y, *et al.* Deep learning vs. Radiomics for predicting axillary lymph node metastasis of breast cancer using ultrasound images: don't forget the peritumoral region. *Front Oncol* 2020; 10: 53.
30. Ning Z, Luo J, Li Y, *et al.* Pattern classification for gastrointestinal stromal tumors by integration of radiomics and deep convolutional features. *IEEE J Biomed Health Inform* 2019; 23: 1181–1191.