

A Multimodal Multipath Artificial Intelligence System for Diagnosing Gastric Protruded Lesions on Endoscopy and Endoscopic Ultrasonography Images

Chang Zhu, PhD^{1,*}, Yifei Hua, MM^{1,*}, Min Zhang, PhD¹, Yun Wang, PhD¹, Wenjie Li, MD¹, Yanbing Ding, MD², Qiang She, MD², Weifeng Zhang, MD¹, Xinmin Si, MD¹, Zihao Kong, MM¹, Baiyun Liu, MD³, Weidao Chen, MD³, Jiangfen Wu, MD³, Yini Dang, MD¹ and Guoxin Zhang, PhD¹

INTRODUCTION: The aim of this study was to develop a novel artificial intelligence (AI) system that can automatically detect and classify protruded gastric lesions and help address the challenges of diagnostic accuracy and inter-reader variability encountered in routine diagnostic workflow.

METHODS: We analyzed data from 1,366 participants who underwent gastroscopy at Jiangsu Provincial People's Hospital and Yangzhou First People's Hospital between December 2010 and December 2020. These patients were diagnosed with submucosal tumors (SMTs) including gastric stromal tumors (GISTs), gastric leiomyomas (GILs), and gastric ectopic pancreas (GEP). We trained and validated a multimodal, multipath AI system (MMP-AI) using the data set. We assessed the diagnostic performance of the proposed AI system using the area under the receiver-operating characteristic curve (AUC) and compared its performance with that of endoscopists with more than 5 years of experience in endoscopic diagnosis.

RESULTS: In the ternary classification task among subtypes of SMTs using modality images, MMP-AI achieved the highest AUCs of 0.896, 0.890, and 0.999 for classifying GIST, GIL, and GEP, respectively. The performance of the model was verified using both external and internal longitudinal data sets. Compared with endoscopists, MMP-AI achieved higher recognition accuracy for SMTs.

DISCUSSION: We developed a system called MMP-AI to identify protruding benign gastric lesions. This system can be used not only for white-light endoscope image recognition but also for endoscopic ultrasonography image analysis.

SUPPLEMENTARY MATERIAL accompanies this paper at <http://links.lww.com/CTG/A893>

Clinical and Translational Gastroenterology 2023;14:e00551. <https://doi.org/10.14309/ctg.0000000000000551>

INTRODUCTION

Benign protruding lesions in the stomach account for approximately 62.5% of upper gastrointestinal lesions and can be classified into epithelial and submucosal lesions (1,2). Although most of these lesions are benign, some protruded lesions, such as gastric stromal tumors (GISTs) and some gastric ectopic pancreas (GEP), have the potential to become malignant phenotypes (3,4). Therefore, a proper diagnosis at the initial clinical evaluation is critical for subsequent decision making (5).

During clinical practice, endoscopists often use a white-light endoscope (WLE) as the initial examination tool to differentiate gastric polyps from submucosal tumors (SMTs) (6). It has been shown that GEP can be easily detected by routine WLE because of

its morphology (7,8), but other subtypes of SMTs remain difficult to differentiate because of their similar visual appearance and subtle differences (9). Endoscopic ultrasonography (EUS), an advanced imaging tool, is considered as a supplementary modality for accurate evaluation of SMTs in the gastrointestinal tract (2,10,11) because it can delineate individual histologic layers and define the site most relevant to the tumor origin (12,13). Moreover, endoscopic ultrasound-guided fine-needle biopsy allows tissue diagnosis to guide further management (5). Nevertheless, diagnosis based on EUS findings requires extensive training and experience of endoscopists, and inconsistent reading and interpretation between endoscopists remains a major clinical challenge (14,15).

¹Department of Gastroenterology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China; ²Department of Gastroenterology, The Affiliated Hospital of Yangzhou University, Yangzhou, China; ³Intervention, Beijing, China. **Correspondence:** Guoxin Zhang, PhD. E-mail: guoxinz@njmu.edu.cn. Yini Dang, MD. E-mail: yeani_hi@126.com.

*Chang Zhu and Yifei Hua contributed equally to this work.

Received April 29, 2022; accepted November 11, 2022; published online November 26, 2022

© 2023 The Author(s). Published by Wolters Kluwer Health, Inc. on behalf of The American College of Gastroenterology

Recent advances in convolutional neural networks (CNNs) have shown remarkable progress in screening medical images (15). Previous studies have demonstrated promising applications of artificial intelligence (AI) in the detection of malignant tumors, such as GI cancer diagnosis, determination of tumor margin, and prediction of invasion depth (16–18). Our research team previously developed a CNN model to detect esophageal protuberant lesions (19), but stomach diseases are much more complex than esophageal diseases. Therefore, we aimed to propose a novel multimodal, multipath AI system (MMP-AI) and evaluate its applicability for the classification of gastric benign protruded lesions.

METHODS

Study design

A total of 1,366 patients who underwent WLE (GIF-Q260J, Olympus, Tokyo, Japan; GIF-Q290J, Olympus) and/or EUS (P2620, FUJI, Chiryu, Japan) between December 2010 and December 2020 were enrolled. The clinical diagnosis of benign gastric protrusion lesions was confirmed based on the pathology records. This retrospective study was approved by the Ethics Committee of the First Affiliated Hospital of Nanjing Medical University (Institutional Review Board No. 2021-SR-023). The need for consent was waived owing to the retrospective nature of this study. All authors had access to the research data and have reviewed and approved the final manuscript.

To imitate realistic clinical practice procedures, our deep learning method was developed as follows (Figure 1):

Training part was composed of 2 sections:

Section 1: Distinguishing subtypes of SMT using a multipath system (either WLE or EUS).

Section 2: Ensembling the individual path in section 1 using a hybrid system.

The validation part comprised internal (Jiangsu Provincial People's Hospital [JSPH]) and external (Yangzhou First People's Hospital [YZPH]) validation.

Data set preparation and image annotations

All endoscopic images were exported as a Joint Photographic Experts Group from the electronic endoscope medical image system, and the quality was reviewed by 2 experienced endoscopists with 5 years of experience (C.Z. and Y.F.H.). We excluded unqualified images based on the following criteria: (i) compromised image quality, including but not limited to blurring, noise, apparent mucus, foam, or food residues that affect the diagnosis visually; (ii) images of other unrelated lesions; and (iii) duplicate images.

Two endoscopists (C.Z. and Y.F.H.) were asked to outline the boundaries of any protruded lesions present in the image, along with the disease label based on the whole image. These masks and image labels were subsequently reviewed by experienced endoscopists (Y.W. and G.X.Z.). Finally, distributions of data sets in each procedure are summarized in Supplementary Digital Content 9, Supplementary Table 1 (<http://links.lww.com/CTG/A893>).

Construction of MMP-AI

We developed a system called MMP-AI to identify gastric benign protruded lesions using endoscopic images. Detailed image

preprocessing and data augmentation are provided in Supplementary Digital Content 1 (<http://links.lww.com/CTG/A893>). In this study, we first implemented CNNs based on ResNeXt, a 50-layer ResNet structure (20), for the classification of gastric benign protrusions on WLE or EUS images (subnetworks 1 and 2). Detailed model architecture and descriptions are provided in Supplementary Digital Content 2 (<http://links.lww.com/CTG/A893>).

Furthermore, for ensembling multimodality gastroscopy image interpretations, we applied the integrated deep convolutional neural networks (DCNN)-long short-term memory network (LSTM) model to discriminate subtypes of gastric submucosal lesions on WLE and EUS images. As illustrated in subnetwork 3, an attention-based bidirectional LSTM model was used to model the temporal interactions between images within WLE/EUS or across WLE/EUS. The detailed structure and description of the LSTM model are presented in Supplementary Digital Content 3 (<http://links.lww.com/CTG/A893>).

Observation experiments

Twelve endoscopists who were not involved in annotating the images participated in the observation experiment: 4 experts (abundant EUS experience with over 5,000 gastroscopies), 4 seniors (basic EUS experience with over 1,000 gastroscopies), and 4 novices (inadequate EUS experience with fewer than 1,000 gastroscopies). All endoscopists were blinded to patients' clinical information and biopsy results. Each endoscopist independently evaluated the digital WLE or EUS images and classified the subtypes of protruded gastric lesions in the images.

Statistical analysis

All statistical analyses were performed using SPSS (version 22.0; SPSS, Chicago, IL) and MedCalc version 15.0 for Windows (MedCalc Software, Ostend, Belgium). Baseline clinical and demographic characteristics were presented as mean \pm SD. The DCNN models from multiple paths were assessed using the area under the receiver-operating characteristic (ROC) curve (AUROC), and 95% confidence intervals (CIs) were determined following the DeLong method. Thereafter, confusion matrices were drawn, and the overall accuracy, sensitivity, specificity, negative predictive value (NPV), positive predictive value (PPV), and F1 score of each target disease category were calculated. We generated the gradient-weighted class activation map to visualize the feature extraction as heatmaps. To compare the model's performance against that of the average endoscopists, we performed the summary ROC (SROC) analysis across all enrolled endoscopists. Statistical significance was set at $P < 0.05$.

RESULTS

Basic characteristics

The proposed models were trained and validated on data sets from 2 individual centers: JSPH (2010–2020) and YZPH (2010–2020). Detailed patient information is presented in Supplementary Table 2 (<http://links.lww.com/CTG/A893>).

Section 1 (Subnetwork 1): SMT subtype classification in WLE-path and EUS-path models

We trained WLE-path and EUS-path models in the system that can categorize subtypes of SMTs into GIL, GEP, and GIST. Detailed information is presented in Supplementary Digital Content 9, Supplementary Table 4 (<http://links.lww.com/CTG/A893>).

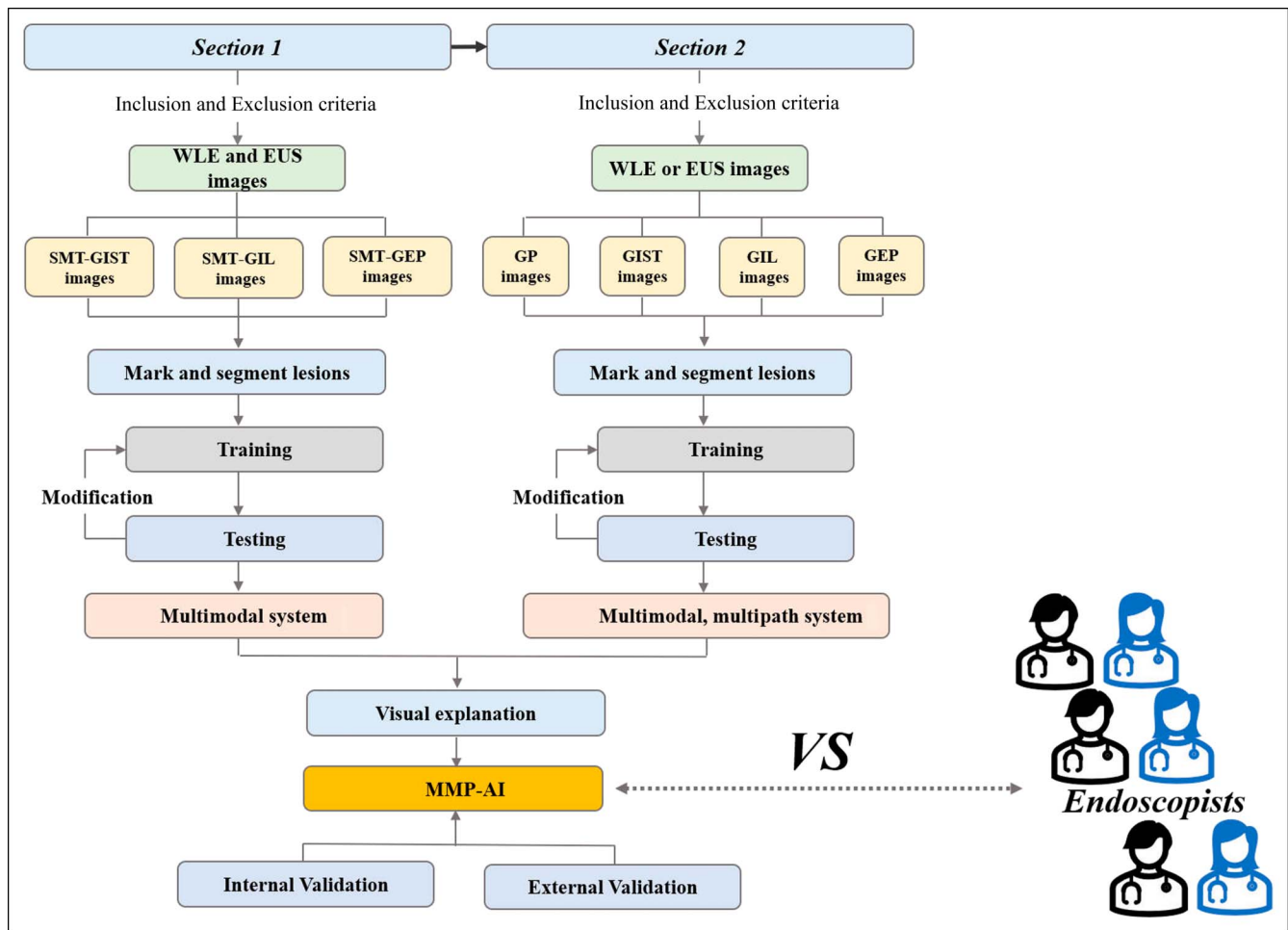


Figure 1. The flowchart of the study design. The study consisted of 2 sections. The construction of MMP-AI procedure included images for training, testing, and validation. Finally, the performance of MMP-AI compared with endoscopists. EUS, endoscopic ultrasonography; GEP, gastric ectopic pancreas; GIL, gastric leiomyoma; GIST, gastric stromal tumor; MMP-AI, multimodal, multipath artificial intelligence system; SMT, submucosal tumor; WLE, white-light endoscopy.

However, there were differences between the prediction results of the 2 models, as shown in the confusion matrix (Figure 2).

Section 2 (Subnetwork 2): SMT subtype classification in the ensemble model

To mimic the realistic clinical diagnosis workflow that endoscopists often refer to as multiple modalities for the accurate identification of SMTs, we merged features learned from the unimodality model path and assembled them into a hybrid model implemented with LSTM. In this way, whenever patients presented with any single modality (WLE or EUS images), the system was able to make a prediction based on the use case. Meanwhile, for patients with paired WLE and EUS images, the system can make further predictions based on the correlated information between these 2 modalities in one patient. This is achieved by the integrated bidirectional LSTM model in subnetwork 2 (details are explained in methods).

As demonstrated in Figure 2, for the SMT classification tasks, the hybrid model achieved area under the curves (AUCs) of 0.890, 0.999, and 0.896 for GIL, GEP, and GIST, respectively. We also calculated the overall accuracy, sensitivity, specificity, and other measurement metrics (Table 1).

Internal validation of the hybrid model in SMT subtype classification

We randomly selected 352 patients from our hospital between 2010 and 2015 as internal validation data set II. Gastroscopic images of these patients were processed using the MMP-AI system. For SMT subtype classification, the hybrid model achieved the highest AUC value in GIST (AUC, 0.913; 95% CI, 0.881–0.942) and GIL (AUC, 0.904; 95% CI, 0.857–0.942). Detailed information is presented in Table 2.

Performance of MMP-AI vs 12 endoscopists

To compare the diagnostic efficacy of the proposed model and endoscopists in protruded gastric lesions, an observational experiment was conducted by 12 endoscopists (different from the endoscopists who annotated the images), including 4 experts, 4 seniors, and 4 novices. The sensitivities and specificities of the endoscopists were plotted on the ROC curve of the trained model, and the average performance of all 12 endoscopists was analyzed and presented as an SROC curve, a meta-analysis reflecting the diagnostic performance (21). We evaluated the classification performance of human endoscopists with different years of experience in predicting the subclass of SMT. The results shown in

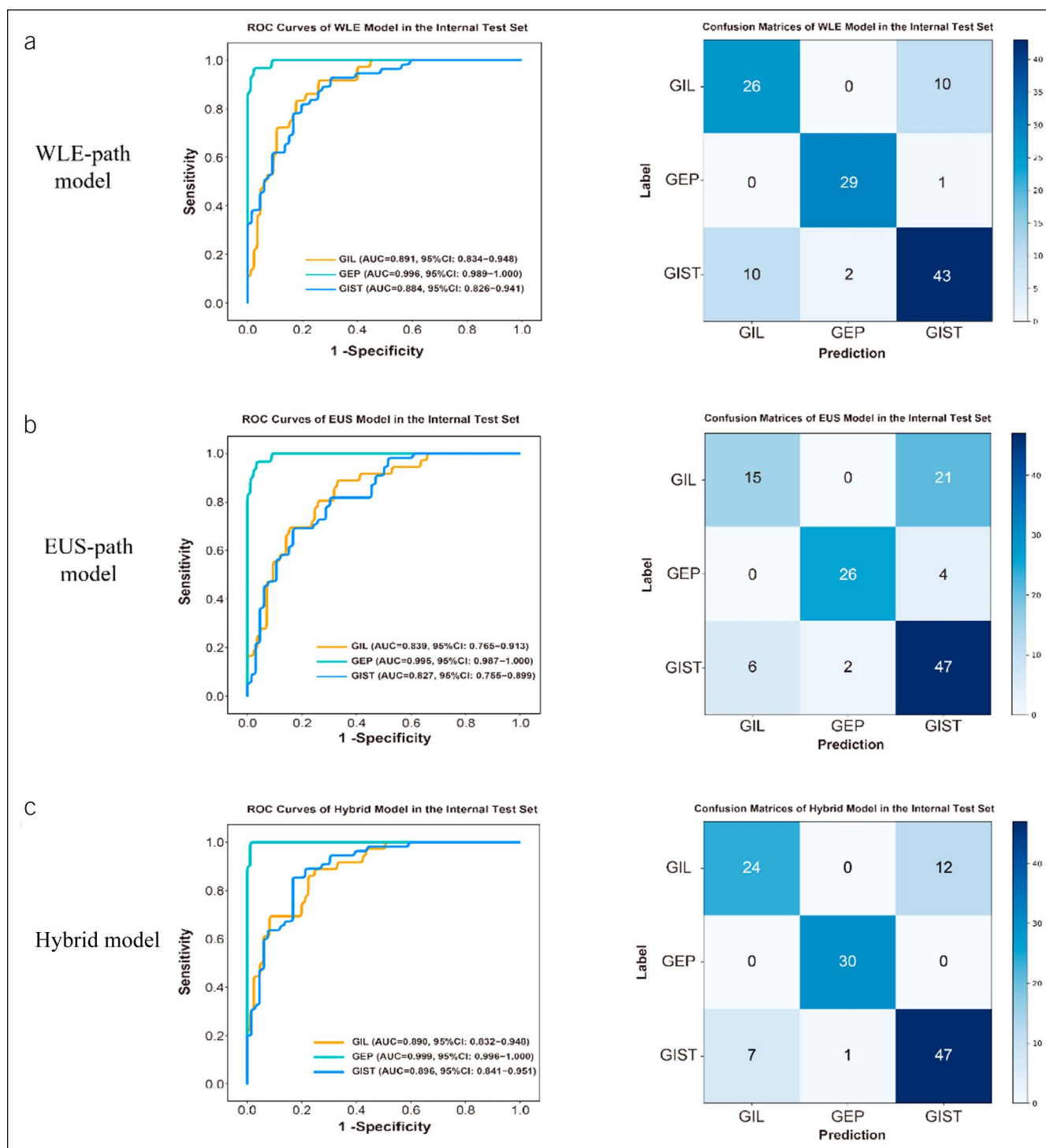


Figure 2. Performance of the multipath system to distinguish subtypes of submucosal tumors (SMTs). (a) Receiver-operating characteristic (ROC) curves and confusion matrix of the white-light endoscope (WLE)-path model. (b) ROC curves and confusion matrix of the endoscopic ultrasonography (EUS)-path model. (c) ROC curves and confusion matrix of the hybrid model. AUC, area under curve; CI, confidence interval; GEP, gastric ectopic pancreas; GIL, gastric leiomyoma; GIST, gastric stromal tumor.

Figure 3 suggest that the hybrid model surpasses endoscopist-level classification performance by showing the area under the ROC curve (Figure 3). When comparing the overall accuracy between the hybrid model and the average endoscopists, the values were 83.50% vs 70.94% for GIST, 78.50% vs 71.76% for GIL, and 98.30% vs 97.66% for GEP. Other classification metrics,

such as sensitivity, specificity, PPV, and NPV, are presented in Table 2.

External verification for MMP-AI

A mature clinical AI system should maintain good performance on other individual data sets. To examine the generalizability of

Table 1. Performance measurements of the hybrid model in internal testing data set I, internal validation data set II, and external validation data set

	AUROC	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	F1 score	AUPRC
Internal testing data set I								
GIL vs others	0.890	78.50	86.10	75.30	59.60	92.80	0.768	0.792
GEP vs others	0.999	98.30	96.70	98.90	96.70	98.90	0.978	0.997
GIST vs others	0.896	83.50	83.60	83.30	80.70	85.90	0.834	0.867
Internal validation data set II								
GIL vs others	0.904	86.80	74.30	89.90	64.20	93.50	0.803	0.760
GEP vs others	0.959	95.80	89.30	98.40	95.80	95.80	0.948	0.952
GIST vs others	0.913	86.60	93.50	79.20	82.70	91.90	0.864	0.901
External validation data set								
GIST vs others	0.785	72.30	80.20	65.30	67.40	78.60	0.723	0.735
GIL vs others	0.736	74.00	59.10	75.60	20.00	94.70	0.570	0.190
GEP vs others	0.903	83.80	72.50	92.50	88.10	81.50	0.831	0.904

AUROC, area under the receiver-operating characteristic curve; AUPRC, area under the precision-recall curve; GEP, gastric ectopic pancreas; GIL, gastrointestinal leiomyoma; GIST, gastrointestinal stromal tumor; NPV, negative predictive value; PPV, positive predictive value.

our MMP-AI, we applied the model to an external validation data set from a different hospital (YZPH) and compared the performance of the model with that of endoscopists.

In the external validation data set, MMP-AI achieved the highest AUC of 0.785 (95% CI, 0.883–0.945) in correctly classifying GISTs compared with GILs (AUC, 0.736; 95% CI, 0.864–0.944) and GEP (AUC, 0.903; 95% CI, 0.931–0.987) (Figure 4). The results shown in Figure 4 suggest that the hybrid model surpasses endoscopist-level classification performance by showing the area under the ROC curve. When comparing the overall accuracy between the hybrid model and the average endoscopists, the values were 72.30% vs 63.30% for GIST, 74.00% vs 70.53% for GIL, and 83.80% vs 73.60% for GEP. Other classification metrics, such as sensitivity, specificity, PPV, and NPV, are provided in Table 2. We also compared the performance of the MMP-AI system with that of 4 endoscopists to make it more clinically relevant. It indicated that the hybrid model surpasses endoscopists' classification performance as shown in the ROC curve (Supplementary Figure 3, <http://links.lww.com/CTG/A893>).

A free-access website was developed to test our MMP-AI system (<https://infer-stomach.infervision.com/home/image>). Any investigator could upload WLE or EUS images to perform this diagnostic system online (Supplementary Digital Content 5, <http://links.lww.com/CTG/A893>).

DISCUSSION

In this study, we developed an MMP-AI system for the automated detection of common gastric protrusions solely from WLE images ($n = 6,406$) or in combination with EUS images ($n = 6,314$). The proposed model was validated using internal longitudinal and external validation data sets, which demonstrated a high degree of classification accuracy and outperformed 12 endoscopists at different experience levels. Particularly, this AI-derived system held diagnostic capacities not only for routine WLE images but also for independent or combined EUS images. These findings have proven robustness in differentiating common subtypes of

protruded lesions with the desired classification performance (GIST, 83.50%; GIL, 78.50%; and GEP, 98.30%). The results raised the possibility of AI-based tools in stratifying multimodal images encountered in clinical practice. In addition, an open-access website was used to test our model online.

Although deep learning-based algorithms have been widely explored in the gastrointestinal field, most studies have focused on malignant tumors, and little attention has been given to benign gastric lesions. For example, Laddha et al (22) reported a real-time YOLOv3 model for gastric polyp detection while its differential diagnostic capacity from other gastric protuberant lesions is under investigation. Kim et al (23) and Minoda et al (24) also developed CNN systems to detect GISTs based on EUS images, with a reported sensitivity and specificity of 83.0% and 75.5%, respectively, in an independent test data set. However, both DL models focused on binary classification tasks and were based solely on EUS images. Early detection and stratification of benign gastric protrusion lesions on multimodal images remain challenging but have a great potential for accurate diagnosis and target interventions. Therefore, this study aimed to develop an MMP-AI-based system that is capable of automated detection of common benign protuberant lesions and prediction of the subtypes of SMTs (GIST, GIL, and GEP) using either WLE or EUS images based on the use case. Notably, sensitivity, specificity, and accuracy of our CNN system for the diagnosis of gastric stromal tumors were better than those of tested by Kim et al in the internal verification and sequential verification (specificity: 83.3% vs 78.0%; sensitivity: 91.0% vs 79.0%; accuracy: 83.5% vs 78.5%), and the sensitivity, specificity, and accuracy were similar to the results observed by Kim in the external verification (specificity: 63.4 vs 78.0%; sensitivity: 80.2% vs 79.0%; accuracy: 72.3% vs 78.5%) (23). In this regard, MMP-AI could be viewed as a powerful addition to assist endoscopists and has the potential to be implemented in the current workflow for the diagnosis of protrusion lesions.

Our MMP-AI system shows value in complex clinical settings and real-world applications. Two internal validation data sets

Table 2. Performance of multipath artificial intelligence system vs 12 endoscopists in internal validation data set II

	Integrated model	Endoscopists			
		Average	Experts	Seniors	Novices
GIL					
Accuracy (%)	78.50	71.76 ^a	74.59	71.49	69.21
Sensitivity (%)	86.10	23.84 ^a	28.47	20.83	22.22
Specificity (%)	75.30	92.06	94.12	92.94	89.12
PPV (%)	59.60	60.65	75.05	60.03	42.28
NPV (%)	92.80	74.69 ^a	75.95	74.09	74.03
F1	0.768	0.279 ^a	0.365	0.246	0.225
GEP					
Accuracy (%)	98.30	97.66 ^a	98.14	97.31	97.52
Sensitivity (%)	96.70	93.89 ^a	95.00	93.33	93.33
Specificity (%)	98.90	98.90	99.18	98.63	98.90
PPV (%)	96.70	96.74	97.58	95.83	96.82
NPV (%)	98.90	98.03 ^a	98.38	97.84	97.86
F1	0.978	0.952 ^a	0.962	0.945	0.949
GIST					
Accuracy (%)	83.50	70.94 ^a	74.38	69.63	68.87
Sensitivity (%)	83.60	87.58	91.36	87.73	83.64
Specificity (%)	83.30	57.07 ^a	60.23	54.55	56.44
PPV (%)	80.70	63.87 ^a	66.15	62.60	62.85
NPV (%)	85.90	87.32	90.21	87.62	84.13
F1	0.834	0.730 ^a	0.765	0.721	0.704

GEP, gastric ectopic pancreas; GIL, gastrointestinal leiomyoma; GIST, gastrointestinal stromal tumor; NPV, negative predictive value; PPV, positive predictive value.
^aSignificant differences compared with the integrated model ($P < 0.05$).

from different time series and 1 external validation data set were used to assess diagnostic robustness and generalizability. Performance metrics in both internal data sets (2015–2020, JSPH and 2010–2015, JSPH) showed excellent performance with an accuracy of 86.60%, 86.80%, and 95.80% and an AUC value of 0.913, 0.904, and 0.959 for the detection of GIST, GIL, and GEP, respectively. The versatility of the system was further proven when using the external validation data set (2010–2015, YZPH),

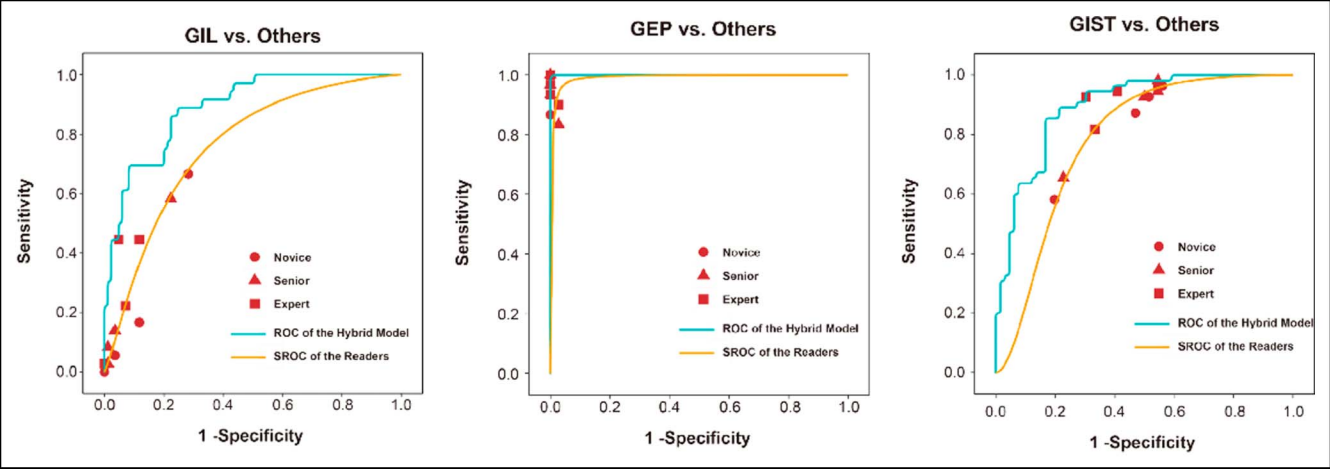


Figure 3. Performance comparisons between the MMP-AI system and 12 endoscopists in internal validation data set II. GEP, gastric ectopic pancreas; GIL, gastric leiomyoma; GIST, gastric stromal tumor; MMP-AI, multimodal, multipath artificial intelligence system; ROC, receiver-operating characteristic; SROC, summary ROC.

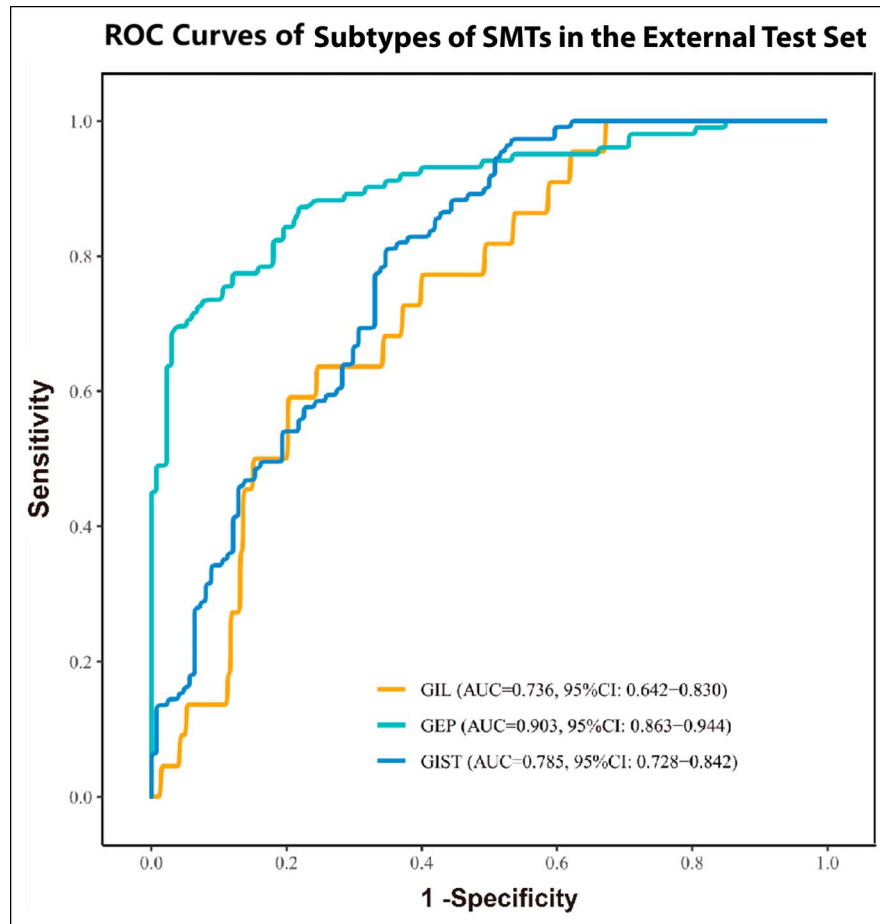


Figure 4. Multicategory classification performance using the proposed MMP-AI model on the external validation data set. AUC, area under the ROC curve; GEP, gastric ectopic pancreas; GIL, gastric leiomyoma; GIST, gastric stromal tumor; MMP-AI, multimodal, multipath artificial intelligence system; ROC, receiver-operating characteristic; SMT, submucosal tumor.

which reflects the various practical applications in real clinical scenarios. Moreover, we used the SROC curve to assess the efficiency of the MMP-AI and endoscopists. This study involved 12 endoscopists from 5 hospitals to reduce selection bias. The diagnostic capability of the MMP-AI was comparable with that of experienced endoscopists (number comparisons). Because the cultivation of experienced endoscopists often requires time and experience, the availability of a cloud-based AI diagnostic system allows fast and accurate endoscopy examination-based diagnosis in primary care settings. In addition, this approach could broaden access, irrespective of regional resource variations between urban and rural areas, leading to improved patient outcomes and potential cost savings.

The innovative design of algorithms is important for the success of our deep learning model. In this study, we innovatively applied attention-based ResNet50 and LSTM networks based on the characteristics of this heterogeneous data set. The ResNet network was used to effectively extract the features of the WLE and EUS images, which ensured the high accuracy of the algorithm in the single-modality mode. Meanwhile, to improve the usefulness and robustness of the model, we further used merged features of the WLE and EUS images to establish the LSTM model for sequential learning, which constructed the relationship between multimodal image features and focused on more valuable image representations to produce accurate predictions.

However, this study has several limitations. First, this was a retrospective study, and only static images were used for model training. The robust performance of the MMP-AI cannot reflect real-time clinical applications. We designed a prospective trial to modify and validate the DCNN system in real-world clinical settings. Second, the sample size for each type of lesion was uneven, particularly the small number of GIL cases, which is related to its low incidence. Third, the MMP-AI system was trained and validated on images obtained using Olympus devices, which might restrain the use of other brands (e.g., Fuji and Pentax). In future studies, we will continue to collect more images using other devices.

We developed an efficient MMP-AI system based on WLE and EUS images that can help doctors identify protruded gastric lesions, circumventing the problem of a lack of EUS in less developed areas.

CONFLICTS OF INTEREST

Guarantor of the article: Guoxin Zhang, PhD, and Yini Dang, MD.
Specific author contributions: G.Z., Y.D., and C.Z.: study concept and design. C.Z., Y.H., M.Z., Y.W., W.L., Y.D., Q.S., W.Z., X.S., and Z.K.: acquisition of data and technique support. C.Z., Y.H., M.Z., B.L., W.C., and J.W.: analysis and interpretation of data. C.Z. and Y.H.: drafting of the manuscript. G.Z. and Y.D.: critical revision of the manuscript.

Financial support: This investigator-initiated trial received no commercial support and has been supported by the grants from the National Natural Science Foundation of China (8197031844), the Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX20_1404), and the medical innovation team of Jiangsu Province of China (No. CXTDA2017033).

Potential competing interests: None to report.

IRB approval statement: This study was approved by the Ethics Committee of the First Affiliated Hospital of Nanjing Medical University, Jiangsu Province Hospital (ethics approval number 2021-SR-023).

ACKNOWLEDGMENTS

The authors thank the following physicians for their efforts: Bixing Ye, Xiaobing Xu, Lurong Li, Ding Heng, Jie Sha, Liping Cheng, Yueyang Gu, Zhi Yang, Jin Zhao, Liangfeng Xu, Huanyu Zhang, and Ziting Miu.

Study Highlights

WHAT IS KNOWN

- ✓ The diagnosis of gastric benign protruded lesions remains a challenging task.
- ✓ The data processing ability of artificial intelligence (AI) has been applied to diagnostic imaging in endoscopy fields.
- ✓ There are few studies of AI focusing on gastric benign protruded lesions.

WHAT IS NEW HERE

- ✓ We developed a system called multipath AI system (MMP-AI) to identify gastric benign protruded lesions
- ✓ MMP-AI held diagnostic capacities not only to routine white-light endoscope images but also to independent or combined endoscopic ultrasonography images.
- ✓ MMP-AI demonstrated a high degree of classification accuracy and outperformed endoscopists.
- ✓ An open-access website has been published to test our model online.

REFERENCES

1. El Chafic AH, Loren D, Siddiqui A, et al. Comparison of FNA and fine-needle biopsy for EUS-guided sampling of suspected GI stromal tumors. *Gastrointest Endosc* 2017;86:510–5.
2. Polkowski M, Butruk E. Submucosal lesions. *Gastrointest Endosc Clinics* 2005;15:33–54.
3. Jeong HY, Yang HW, Seo SW, et al. Adenocarcinoma arising from an ectopic pancreas in the stomach. *Endoscopy* 2002;34:1014–7.
4. Cazacu IM, Luzuriaga Chavez AA, Nogueras Gonzalez GM, et al. Malignant transformation of ectopic pancreas. *Dig Dis Sci* 2019;64:655–68.
5. Polkowski M. Endoscopic ultrasound and endoscopic ultrasound-guided fine-needle biopsy for the diagnosis of malignant submucosal tumors. *Endoscopy* 2005;37:635–45.
6. Morais DJ, Yamanaka A, Zeitune JMR, et al. Gastric polyps: A retrospective analysis of 26,000 digestive endoscopies. *Arq Gastroenterol* 2007;44:14–7.
7. Hedenbro JL, Ekelund M, Wetterberg P. Endoscopic diagnosis of submucosal gastric lesions. *Surg Endosc* 1991;5:20–3.
8. Thoeni RF, Gedgudas RK. Ectopic pancreas: Usual and unusual features. *Gastrointest Radiol* 1980;5:37–42.
9. Perrillo RP, Zuckerman GR, Shatz BA. Aberrant pancreas and leiomyoma of the stomach: Indistinguishable radiologic and endoscopic features. *Gastrointest Endosc* 1977;23:162–3.
10. Landi B, Palazzo L. The role of endosonography in submucosal tumours. *Best Pract Res Clin Gastroenterol* 2009;23:679–701.
11. Hwang JH, Saunders MD, Rulyak SJ, et al. A prospective study comparing endoscopy and EUS in the evaluation of GI subepithelial masses. *Gastrointest Endosc* 2005;62:202–8.
12. Papanikolaou IS, Triantafyllou K, Kourikou A, et al. Endoscopic ultrasonography for gastric submucosal lesions. *World J Gastrointest Endosc* 2011;3:86.
13. Okten RS, Kacar S, Kucukay F, et al. Gastric subepithelial masses: Evaluation of multidetector CT (multiplanar reconstruction and virtual gastroscopy) versus endoscopic ultrasonography. *Abdom Imaging* 2012;37:519–30.
14. Rösch T. State of the art lecture: Endoscopic ultrasonography: Training and competence. *Endoscopy* 2006;39:69–72.
15. Van Dam J, Brady P, Freeman M, et al. Guidelines for training in electronic ultrasound: Guidelines for clinical application. From the ASGE. American Society for Gastrointestinal Endoscopy. *Gastrointest Endosc* 1999;49:829–33.
16. Luo H, Xu G, Li C, et al. Real-time artificial intelligence for detection of upper gastrointestinal cancer by endoscopy: A multicentre, case-control, diagnostic study. *Lancet Oncol* 2019;20:1645–54.
17. Namikawa K, Hirasawa T, Nakano K, et al. Artificial intelligence-based diagnostic system classifying gastric cancers and ulcers: Comparison between the original and newly developed systems. *Endoscopy* 2020;52:1077–83.
18. Tang D, Wang L, Ling T, et al. Development and validation of a real-time artificial intelligence-assisted system for detecting early gastric cancer: A multicentre retrospective diagnostic study. *EBioMedicine* 2020;62:103146.
19. Zhang M, Zhu C, Wang Y, et al. Differential diagnosis for esophageal protruded lesions using a deep convolution neural network in endoscopic images. *Gastrointest Endosc* 2021;93:1261–72.e2.
20. Zhang H, Wu C, Zhang Z, et al. ResNest: Split-attention networks. 2020. arXiv preprint arXiv:2004.08955.
21. Oakden-Rayner L, Palmer L. Docs are ROCs: A simple off-the-shelf approach for estimating average human performance in diagnostic studies. 2020. arXiv preprint arXiv:2009.11060.
22. Laddha M, Jindal S, Wojciechowski J. Gastric polyp detection using deep convolutional neural network. In: *Proceedings of the 2019 4th International Conference on Biomedical Imaging, Signal Processing*; 2019. p. 55–9.
23. 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.
24. 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;55:1119–26.

Open Access This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.