



Published in final edited form as:

Neurogastroenterol Motil. 2023 July ; 35(7): e14549. doi:10.1111/nmo.14549.

An artificial intelligence platform provides an accurate interpretation of esophageal motility from Functional Lumen Imaging Probe Panometry studies

Wenjun Kou¹, Priyanka Soni², Matthew W. Klug³, Mozziyar Etemadi^{2,3}, Peter J. Kahrilas¹, John E. Pandolfino¹, Dustin A. Carlson¹

¹Division of Gastroenterology and Hepatology, Department of Medicine, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA

²Department of Anesthesiology, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA

³Department of Information Services, Northwestern Medicine, Chicago, Illinois, USA

Abstract

Background: Functional lumen imaging probe (FLIP) Panometry is performed at the time of sedated endoscopy and evaluates esophageal motility in response to distension. This study aimed to develop and test an automated artificial intelligence (AI) platform that could interpret FLIP Panometry studies.

Methods: The study cohort included 678 consecutive patients and 35 asymptomatic controls that completed FLIP Panometry during endoscopy and high-resolution manometry (HRM). “True” study labels for model training and testing were assigned by experienced esophagologists per a hierarchical classification scheme. The supervised, deep learning, AI model generated FLIP Panometry heatmaps from raw FLIP data and based on convolutional neural networks assigned esophageal motility labels using a two-stage prediction model. Model performance was tested on a 15% held-out test set ($n = 103$); the remainder of the studies were utilized for model training ($n = 610$).

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial License](#), which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

Correspondence: Dustin A. Carlson, Feinberg School of Medicine, Department of Medicine, Division of Gastroenterology and Hepatology, Northwestern University, 676 St Clair St, Suite 1400, Chicago, IL 60611-2951, USA. dustin-carlson@northwestern.edu. Wenjun Kou and Priyanka Soni Co-first authors.

AUTHOR CONTRIBUTIONS

WK and PS contributed to drafting of the manuscript, data analysis, data interpretation, and approval of the final version. MWK contributed to data analysis, data interpretation, editing the manuscript critically and approval of the final version. ME contributed to study concept, data analysis, data interpretation, and approval of the final version. PJK contributed to editing the manuscript critically and approval of the final version. JEP contributed to study concept and design, obtaining funding, data interpretation, editing the manuscript critically, and approval of the final version. DAC contributed to study concept and design, data analysis, data interpretation, drafting of the manuscript, obtaining funding, and approval of the final version.

DISCLOSURE

WK, PS, MWK, ME, JEP, PJK, DAC, and Northwestern University hold shared intellectual property rights and a licensing agreement with Medtronic Inc. WK: Bristol-Myers Squibb (Consulting). PJK: AstraZeneca (consulting), Ironwood (Consulting); Reckitt (Consulting), Johnson and Johnson (consulting). JEP: Sandhill Scientific/Diversatek (Consulting, Speaking, Grant), Takeda (Speaking), Astra Zeneca (Speaking), Medtronic (Speaking, Consulting, Patent, License), Torax (Speaking, Consulting), Ironwood (Consulting). DAC: Medtronic (Speaking, Consulting); Phathom Pharmaceuticals (Consulting)

Key Results: “True” FLIP labels across the entire cohort included 190 (27%) “normal,” 265 (37%) “not normal/not achalasia,” and 258 (36%) “achalasia.” On the test set, both the Normal/Not normal and the achalasia/not achalasia models achieved an accuracy of 89% (with 89%/88% recall, 90%/89% precision, respectively). Of 28 patients with achalasia (per HRM) in the test set, 0 were predicted as “normal” and 93% as “achalasia” by the AI model.

Conclusions: An AI platform provided accurate interpretation of FLIP Panometry esophageal motility studies from a single center compared with the impression of experienced FLIP Panometry interpreters. This platform may provide useful clinical decision support for esophageal motility diagnosis from FLIP Panometry studies performed at the time of endoscopy.

Keywords

dysphagia; impedance; machine learning; spasm

1 | INTRODUCTION

An evaluation for esophageal motility disorders is recommended for evaluation of esophageal dysphagia or chest pain when a mechanical esophageal obstruction is not detected on upper endoscopy.¹ While esophageal manometry is the conventional test to diagnosis esophageal motility disorders, functional lumen imaging probe (FLIP) Panometry, represents a novel method to evaluate esophageal motility.^{2–4} We described classifying esophageal motility using FLIP Panometry and demonstrated that the FLIP Panometry motility classifications frequently paralleled the esophageal motility diagnoses provided by high-resolution manometry (HRM) and the Chicago Classification v4.0 (CCv4.0).^{4,5} In particular, a normal FLIP Panometry was associated with a normal (or equivocal) HRM diagnosis in 95% of patients while more than 99% of patients with a manometric diagnosis of achalasia had an abnormal FLIP Panometry study.

In clinical practice, FLIP Panometry and HRM may be used in a complementary manner for esophageal motility diagnosis, particularly as recommended by CCv4.0 when there is an inconclusive HRM diagnosis, such as esophagogastric junction (EGJ) outflow obstruction (EGJO).^{5,6} In some clinical scenarios, FLIP Panometry could potentially be utilized as the primary method for esophageal motility evaluation, such as if FLIP Panometry is normal (essentially ruling out a major esophageal motility disorder), or if a patient is unable to tolerate a transnasal HRM catheter.^{4,7} Because FLIP is performed during endoscopy, it offers an advantage over HRM by measuring esophageal motility comfortably in a sedated patient, as well as providing esophageal motility evaluation concurrently with endoscopy.⁸ While the FLIP Panometry motility classification is based on pattern recognition of contractile response (secondary peristalsis) patterns and EGJ-distensibility metrics, there is a component of subjective interpretation that can sometimes be challenging based on the dynamic aspect of the FLIP study when performed during the endoscopic encounter.

Overall, an automated decision support tool to facilitate interpretation of esophageal motility findings from the FLIP Panometry study is appealing. We recently demonstrated that artificial intelligence (AI) models were able to accurately identify esophageal motility

diagnoses from raw HRM data.^{9,10} This study aimed to develop and test an AI platform that could predict FLIP Panometry motility diagnoses.

2 | METHODS

2.1 | Subjects

Consecutive, adult patients (ages 18–89 years) that underwent evaluation of esophageal symptoms between November 2012 and December 2019 and completed FLIP during upper endoscopy and HRM suitable for CCv4.0 were included (Table 1; Figure 1); this study cohort (patients and controls) have been previously described.^{4,11,12} Patients with previous foregut surgery (including previous pneumatic dilation) or esophageal mechanical obstructions including esophageal stricture, eosinophilic esophagitis, severe reflux esophagitis (Los Angeles-classification C or D), hiatal hernia >3 cm were excluded as these are potential causes of secondary esophageal motor abnormalities and preclude application of CCv4.0 (Figure 1).⁵ Additional baseline clinical evaluation with timed barium esophagram (TBE) was obtained at the discretion of the primary treating gastroenterologist. The study protocol was approved by the Northwestern University Institutional Review Board (STU00210464) as minimal risk with a waiver of informed consent for analysis of deidentified, coded patient data.

2.2 | FLIP study protocol and analysis

The FLIP study using 16-cm FLIP (EndoFLIP® EF-322 N; Medtronic, Inc, Shoreview, MN) was performed during sedated endoscopy as previously described.^{3,13,14} The FLIP study included stepwise 10-mL FLIP distensions (each stepwise distension volume being maintained for 30–60 s) with the FLIP catheter positioned across the EGJ (1–3 intragastric channels). The manual FLIP Panometry analysis and data labeling was performed remote from endoscopy using a customized program (available open source at <http://www.wklytics.com/nmgi>) and focused on the 50, 60, and 70 mL fill volumes. The FLIP Panometry analysis was performed blinded to other clinical data, including endoscopy, HRM and TBE results, as previously described and summarized in Table 2.^{4,11,12} The FLIP Panometry labels were assigned by agreement between experienced raters (DAC and JEP). The contractile response pattern was based on review of esophageal contractility during the 50, 60, and 70 mL fill volumes with specific features and patterns of contractility (such as repetitive antegrade contractions or sustained occluding contractions) that were then applied to assign a contractile response (CR) pattern.^{4,11} EGJ opening was classified by applying the EGJ-distensibility index (DI) at the 60 mL FLIP fill volume and the maximum EGJ diameter that was achieved during the 60 mL or 70 mL fill volume.^{4,15} The contractile response pattern and EGJ opening classification were then applied to assign a FLIP Panometry motility classification (Table 2).^{4,15} The FLIP Panometry studies were then classified as “normal” or “not normal” and then the not normal studies as suspected “achalasia” or “not achalasia” based on previous findings evaluating the association of FLIP Panometry motility with HRM/CCv4.0 diagnoses, Figure 2.⁴

2.3 | FLIP Panometry motility interpretation models

For the AI models, the data were processed by taking the raw diameter readings from the FLIP study and transforming it into a FLIP Panometry heatmap (Figure 3), which allowed utilization of convolutional neural networks/layers for image inspection.

Based on how the FLIP Panometry studies can be pragmatically applied in clinical practice, a two-stage model pipeline was developed (labels per Figure 2). In stage 1, the model predicted as “normal” versus “not normal.” If the model found the study to be “not normal,” it was further classified in Stage 2 through a suspected “achalasia” versus “not achalasia” model.

- Stage (1) “Normal” versus “not normal”: A custom multiheaded convolutional neural network was developed with the goal of each head of the model capturing different patterns identified as important in “Normal” studies. Each of the inputs were passed using various kernel sizes and strides to capture the distinguishing patterns.
- Stage (2) Not normal: suspected “achalasia” versus “not achalasia”: To account for the removal of normal studies from the data, transfer learning/pretraining was leveraged to create a base model to build upon. The VGG16 network pretrained on the ImageNet dataset was selected as our base and then an additional set of linear layers were added for fine-tuning on the Stage 2 labels.

The performance of each of the AI models for the primary analysis was tested on a 15% held-out test set ($n = 103$) that utilized stratified sampling to maintain proportionate sampling distribution of each label; the remainder of the study cohort ($n = 610$) was utilized for model training (Figure 1).

Additional clinical factors, including HRM/CCv4.0 motility diagnoses and TBE findings were examined relative to the AI model labeling, with a focus on studies with “inaccurate” predictions from the models.

2.4 | HRM and TBE protocol and analysis

HRM studies and interpretation were completed as previously described using a solid-state assembly with 36 circumferential pressure sensors at 1-cm intervals (Medtronic Inc, Shoreview, MN) according to the CCv4.0.^{4–6} HRM studies were interpreted independent of FLIP results.

Timed barium esophagram (TBE) with barium tablet was obtained in patients at the discretion of the patients’ treating physicians. The barium column height above the EGJ was measured from images obtained at 1, 2 and 5 min after ingestion of 200 mL barium. If liquid barium cleared, a 12.5 mm barium tablet was administered. TBE results were categorized for analysis based on the findings of greatest severity by: (a) 5-min column height >5 cm, (b) 1-min column height >5 cm or impaction of a 12.5 mm barium tablet (i.e., inability of the barium tablet to pass), or (c) “normal” (i.e., not meeting preceding severity criteria).

For studies with an HRM/CCV4.0 classification of EGJOO (i.e., an “inconclusive” HRM diagnosis in isolation), timed barium esophagram (TBE) findings were applied when available. Patients with HRM-EGJOO were further defined as “conclusive EGJOO” when the TBE had either (a) 5-min column height >5 cm or (b) a 1-min column height >5 cm and also impaction of a 12.5 mm barium tablet. Patients were otherwise labeled as “inconclusive EGJOO” with other “inconclusive TBE” findings, or if TBE was not completed.

2.5 | Statistical analysis

To describe the clinical characteristics of the cohort (as well as among specified subgroups), results were reported as n (%), mean (standard deviation; SD), or median (interquartile range; IQR) as appropriate based on the variable type and depending on data distribution. Comparisons of categorical variables were performed between subject subgroups using Chi-square tests. Comparisons of continuous variables were performed between subject subgroups using ANOVA/t-tests for normally distributed variables and using Kruskal-Wallis/Mann-Whitney U for non-normally distributed variables. Statistical significance was considered at a two-tailed p -value <0.05. For comparisons involving significant differences between more than two groups, post-hoc comparison testing was completed using a Bonferroni correction to address multiple comparisons.

The performance of the AI models was tested on a 15% held-out test set ($n = 103$) that utilized stratified sampling to maintain sampling distribution of each label. AI model performance was evaluated on recall, precision, and overall area-under-the-curve metrics on the held-out test set.

3 | RESULTS

3.1 | Subjects

678 patients, mean (SD) age 54 (16) years, 57% female and 35 asymptomatic controls, mean (SD) age 30 (6) years, 71% female were included (Table 1). The majority (90%) of patients were evaluated for dysphagia. Among the entire subject cohort, the “true” FLIP labels were “normal” in 190 (27%) studies, “not normal”/“not achalasia” in 265 (37%) studies, and “not normal”/suspected “achalasia” in 258 (36%) studies. The FLIP Panometry motility classifications included 218 normal (31%; including all 35 controls), 43 (6%) weak, 235 (33%) obstruction with weak contractile response, 77 (11%) spastic-reactive, and 140 (20%) inconclusive. The most common HRM diagnoses were achalasia (227 patients; 32% of the cohort) and normal motility (250 patients and 32/35 controls; 35% of the cohort). 318 (47%) patients completed a TBE.

The training and test cohorts were similar with regard to demographics, true FLIP Panometry motility labels, HRM motility classifications, and proportions of TBE completion and findings (Table 1).

3.2 | Performance of two-stage FLIP prediction model

The normal/not normal model achieved 89% accuracy (95% confidence interval 0.83–0.95) with 89% weighted average recall/sensitivity and 90% weighted average precision on the

held-out test set (Figure 4). All four asymptomatic controls in the held-out test set were predicted as “normal,” whereas there were 0 patients with achalasia (achalasia on HRM or suspected “achalasia” on FLIP Panometry) that the model predicted as “normal;” Table 3.

There were 65 patients from the held-out set with a predicted “not normal” FLIP that were then tested in stage 2, the achalasia/not achalasia model. This model achieved an accuracy of 89% (95% confidence interval 0.81–0.97), 88% weighted average recall/sensitivity, and 89% weighted average precision (Figure 4). Of the 28 patients with HRM/CCv4.0 diagnosis of achalasia in the test cohort, 93% were predicted as suspected achalasia by the model, as were 3/3 patients with a diagnosis of conclusive EGJOO (*conclusive* based on HRM and TBE).

3.3 | Evaluation of “inaccurate” model predictions

“Inaccurate” model predictions from the test cohort inaccurate case are described in Figure 4 and each of the 18 individual studies are detailed in Table 4. Of the 8 studies inaccurately predicted normal, none had a “true” FLIP label of achalasia, nor did any have a HRM/CCv4.0 diagnosis of achalasia (achalasia subtype I, II, III or conclusive EGJOO); Figure 4. 6/8 had a true FLIP motility classification of “weak” (i.e., with an abnormal contractile response and normal EGJ opening), while the remaining 2/8 had an inconclusive FLIP motility classification (both had borderline-normal EGJ opening).

Additionally, none of the 5 studies inaccurately predicted as “suspected achalasia” had a true FLIP label of “normal.” Further, all five such studies had a true FLIP motility classification of “inconclusive” with borderline EGJ opening and abnormal contractile responses (Table 4).

4 | DISCUSSION

The major finding of this study, based on a cohort of 678 esophageal motility patients and 35 healthy controls, was that an AI model accurately interpreted esophageal motility classifications from FLIP Panometry studies with 89% accuracy in both stages of a two-stage model using a simple, clinically pragmatic classification scheme for esophageal motility disorders. Even when the AI model was “inaccurate,” the model interpretation was generally reasonable (e.g., often to adjacent classifications), and furthermore, the “inaccuracy” would typically be associated with minimal clinical consequence (i.e., no patients with achalasia were predicted as “normal”). Overall, this study suggests that an AI model for FLIP Panometry interpretation can provide reliable decision support for FLIP Panometry studies performed clinically during endoscopy.

There has been increasing use of AI and machine learning in medicine and gastroenterology, such as for assistance of visual inspection of endoscopic images for colon polyps or Barrett’s esophagus.^{16,17} The model described here transformed raw FLIP data into FLIP Panometry heatmaps that allowed for image inspection by deep-learning AI models, similar to the those utilized on endoscopy images. The supervised, deep-learning model demonstrated herein utilized convolutional neural networks and careful parameter selection of the convolutional layers to inspect the FLIP studies in a manner somewhat similar

to a clinician. While this study is the first to develop and test an AI model for FLIP Panometry motility interpretation, we have recently described using other AI approaches for HRM interpretation.^{9,18} Similar to using deep-learning algorithms to interpret the graphical depiction of esophageal motility provided by esophageal pressure topography, the models described interpreted esophageal motility from esophageal diameter topography (and pressure) from FLIP Panometry data. With both manometry and FLIP, these AI approaches offer the potential to provide clinical decision support for esophageal motility interpretation.

We previously described using feature-based machine learning on FLIP Panometry metrics to facilitate prediction of HRM-based achalasia subtypes with 90% and 70% accuracy in train and test cohorts, respectively.¹⁹ Recognizing that there are advantages to both image-based deep learning and feature-based AI/machine learning approaches, we demonstrated that a multistage model integrating a balanced combination of deep-learning and feature-based AI/machine learning models accurately predicted motility diagnosis on HRM.¹⁰ Given the quantifiable physiometers provided with FLIP Panometry, an appealing future direction with FLIP Panometry motility interpretation will be to develop hybrid models with a similar framework.

FLIP Panometry done concurrently with endoscopy may be applied in clinical practice such that when a FLIP Panometry study is normal (at either an index endoscopy or following an inconclusive initial evaluation, e.g., EGJOO on HRM), there is a low probability for a major motility disorder. Hence achalasia, the most important diagnosis in an esophageal motility evaluation, is essentially ruled out and management may be directed toward GERD or other syndrome; Figure 5.^{4,7} Given the low probability for a major esophageal motor disorder, manometry may be avoided in this scenario (or at least delayed if symptoms persist after initial treatment). Conversely, when FLIP Panometry is abnormal, the FLIP results can be incorporated with the existing clinical impression based on clinical presentation and endoscopic findings (i.e., a pretest probability) to support (or refute) achalasia. If there is a high pretest probability for achalasia (e.g., supportive clinical history and endoscopy) and a FLIP prediction of achalasia, treatment could potentially be pursued without need for HRM (especially if the patient is unable to tolerate HRM). Though if needed, the clinical motility diagnosis can then be ultimately reached via complementary application of other data, for example, HRM and/or TBE.

With regard to the potential application of the described AI model in clinical practice, it is expected that less experienced FLIP Panometry users may rely more heavily on the model interpretations, while more experienced FLIP Panometry users will use the model interpretations as reassurance when their independent interpretation agrees with the model, or prompt additional scrutiny of the FLIP study when interpretations differ. In either case, the results of this study demonstrated that a “normal” FLIP interpretation resulted in zero patients with achalasia being “missed” and an initial management plan targeting GERD would have been reasonable for all such patients, even the few patients eventually diagnosed with a disorder of primary peristalsis. Further, a prediction for an abnormal FLIP Panometry would likely lead to application of complementary data with TBE or HRM (Figure 5). Notably, all patients with an HRM/CCv4.0 diagnosis of achalasia in the held-out test set were predicted as “not normal” and 93% of these patients had a predicted FLIP diagnosis

of “suspected achalasia.” Overall, application of the model predictions to clinical practice would be expected to facilitate accurate detection (or exclusion) of esophageal motility disorders via FLIP Panometry and provide clinical decision support when interpreting FLIP Panometry studies done concurrently with endoscopy.

There are several strengths of this study investigating novel AI models for FLIP classification including a large clinical cohort (including healthy controls) and comprehensive motility testing independent of FLIP (HRM interpreted per CCv4.0 criteria and a subset with TBE) to complement the clinical relevance of the model predictions. However, the study has limitations as well. While we explored the clinical relevance of the model predictions using additional data independent of FLIP (e.g., HRM and TBE), this at times may be an unfair measure by which to judge the AI model interpretations, noting that there is no singular “gold standard” test for esophageal motility disorders. Establishing an absolute ground truth for esophageal motility disorders can be challenging being that FLIP, HRM, and TBE each have inherent limitations and none are perfect tests. Thus, instead of representing a limitation of model performance, discrepancies may reflect that motility on FLIP and HRM sometimes differ, for example, secondary peristalsis (FLIP) and primary peristalsis (HRM) can differ within individuals.^{11,20,21} Another limitation is that this work describes a single center study, thus a multicenter study is ongoing to test this model on external patient cohorts to further validate this model and demonstrate generalizability. Additionally, future work is needed to help develop models for prediction of targeted treatment outcomes, which likely will involve incorporation of data from complimentary tests. AI/machine learning offers a promising approach to facilitate this multimodal integration.

In conclusion, AI models were able to accurately interpret esophageal motility per a simple, clinically pragmatic classification of esophageal motility disorders using FLIP Panometry studies from a single center suggesting that this technology has significant potential to provide clinical decision support for FLIP Panometry studies done concurrently with endoscopy. Future work is anticipated to refine both FLIP Panometry and clinical motility diagnosis labeling, as well as incorporating longitudinal treatment outcomes, to further develop advanced models for diagnosis of esophageal motility disorders. The promise of AI for clinical decision support is apparent and represents potential for exciting advances in gastroenterology and medicine.

FUNDING INFORMATION

This work was supported by P01 DK117824 (JEP) from the Public Health service, American College of Gastroenterology Junior Faculty Development Award (DAC), and gifts from Joe and Nives Rizza and The Todd and Renee Schilling Charitable Fund.

Funding information

National Institute of Diabetes and Digestive and Kidney Diseases

Abbreviations:

AI	Artificial intelligence
----	-------------------------

CCv4.0	Chicago Classification v4.0
DI	Distensibility index
EGJ	Esophagogastric junction
EGJOO	Esophagogastric junction outflow obstruction
FLIP	Functional lumen imaging probe
HRM	High-resolution manometry
RACs	Repetitive Antegrade Contraction
TBE	Timed barium esophagram

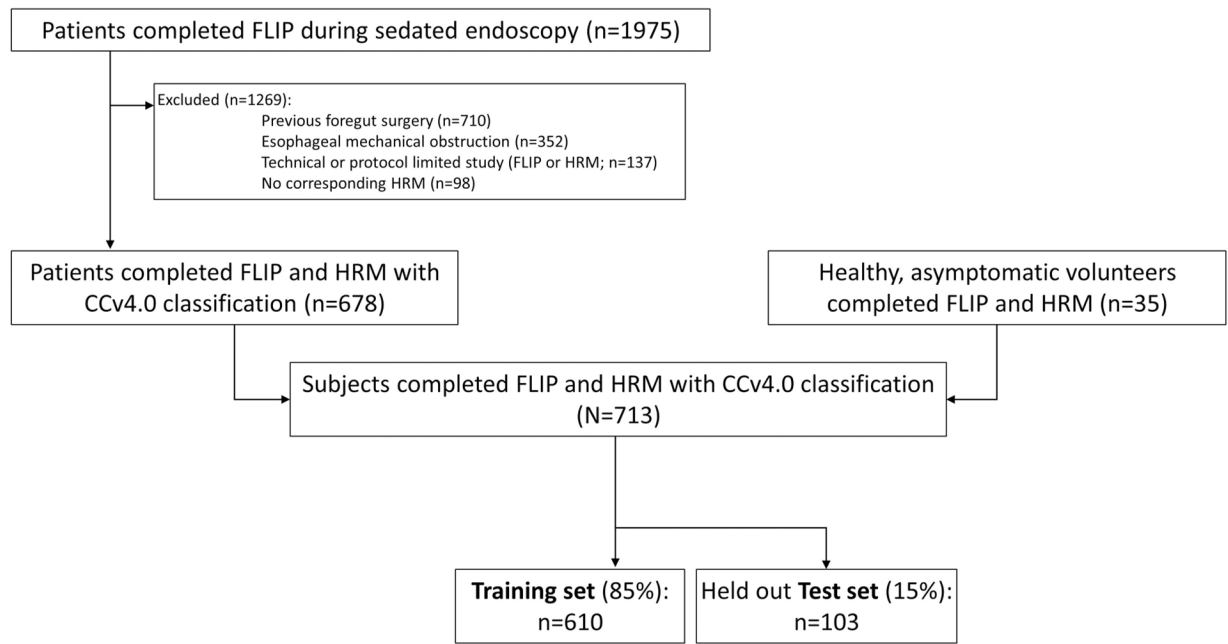
REFERENCES

1. Gyawali CP, Carlson DA, Chen JW, Patel A, Wong RJ, Yadlapati RH. ACG clinical guidelines: clinical use of esophageal physiologic testing. *Am J Gastroenterol*. 2020;115(9):1412–1428. [PubMed: 32769426]
2. Carlson DA, Lin Z, Rogers MC, Lin CY, Kahrilas PJ, Pandolfino JE. Utilizing functional lumen imaging probe topography to evaluate esophageal contractility during volumetric distention: a pilot study. *Neurogastroenterol Motil*. 2015;27(7):981–989. [PubMed: 25898916]
3. Carlson DA, Kahrilas PJ, Lin Z, et al. Evaluation of esophageal motility utilizing the functional lumen imaging probe. *Am J Gastroenterol*. 2016;111(12):1726–1735. [PubMed: 27725650]
4. Carlson DA, Gyawali CP, Khan A, et al. Classifying esophageal motility by FLIP Panometry: a study of 722 subjects with Manometry. *Am J Gastroenterol*. 2021;116(12):2357–2366. [PubMed: 34668487]
5. Yadlapati R, Kahrilas PJ, Fox MR, et al. Esophageal motility disorders on high-resolution manometry: Chicago classification version 4.0(cc). *Neurogastroenterol Motil*. 2021;33(1):e14058. [PubMed: 33373111]
6. Bredenoord AJ, Babaei A, Carlson D, et al. Esophagogastric junction outflow obstruction. *Neurogastroenterol Motil*. 2021;33(9):e14193. doi:10.1111/nmo.14193 [PubMed: 34120375]
7. Baumann AJ, Donnan EN, Triggs JR, et al. Normal functional luminal imaging probe Panometry findings associate with lack of major esophageal motility disorder on high-resolution Manometry. *Clin Gastroenterol Hepatol*. 2021;19(2):259–268.e1. [PubMed: 32205217]
8. Carlson DA, Gyawali CP, Kahrilas PJ, et al. Esophageal motility classification can be established at the time of endoscopy: a study evaluating real-time functional luminal imaging probe panometry. *Gastrointest Endosc*. 2019;90(6):915–923.e1. [PubMed: 31279625]
9. Kou W, Carlson DA, Baumann AJ, et al. A deep-learning-based unsupervised model on esophageal manometry using variational autoencoder. *Artif Intell Med*. 2021;112:102006. [PubMed: 33581826]
10. Kou W, Carlson DA, Baumann AJ, et al. A multi-stage machine learning model for diagnosis of esophageal manometry. *Artif Intell Med*. 2022;124:102233. [PubMed: 35115131]
11. Carlson DA, Baumann AJ, Prescott JE, et al. Validation of secondary peristalsis classification using FLIP panometry in 741 subjects undergoing manometry. *Neurogastroenterol Motil*. 2021;34(1):e14192. doi:10.1111/nmo.14192 [PubMed: 34120383]
12. Carlson DA, Prescott JE, Baumann AJ, et al. Validation of clinically relevant thresholds of Esophagogastric junction obstruction using FLIP Panometry. *Clin Gastroenterol Hepatol*. 2022;20(6):e1250–e1262. doi:10.1111/nmo.14290 [PubMed: 34216821]
13. Carlson DA, Kou W, Lin Z, et al. Normal values of esophageal Distensibility and distension-induced contractility measured by functional luminal imaging probe Panometry. *Clin Gastroenterol Hepatol*. 2019;17(4):674–681.e1. [PubMed: 30081222]

14. Carlson DA, Baumann AJ, Donnan EN, Krause A, Kou W, Pandolfino JE. Evaluating esophageal motility beyond primary peristalsis: assessing esophagogastric junction opening mechanics and secondary peristalsis in patients with normal manometry. *Neurogastroenterol Motil.* 2021;33:e14116. [PubMed: 33705590]
15. Carlson DA, Prescott JE, Baumann AJ, et al. Validation of clinically relevant thresholds of Esophagogastric junction obstruction using FLIP Panometry. *Clin Gastroenterol Hepatol.* 2022;20(6):e1250–e1262. [PubMed: 34216821]
16. Parasa S, Wallace M, Bagci U, et al. Proceedings from the first global artificial intelligence in gastroenterology and endoscopy summit. *Gastrointest Endosc.* 2020;92(4):938–45.e1. [PubMed: 32343978]
17. Rodriguez-Diaz E, Baffy G, Lo WK, et al. Real-time artificial intelligence-based histologic classification of colorectal polyps with augmented visualization. *Gastrointest Endosc.* 2021;93(3):662–670. [PubMed: 32949567]
18. Kou W, Galal GO, Klug MW, et al. Deep learning-based artificial intelligence model for identifying swallow types in esophageal high-resolution manometry. *Neurogastroenterol Motil.* 2022;34(7):e14290. doi:10.1111/nmo.14290 [PubMed: 34709712]
19. Carlson DA, Kou W, Rooney KP, et al. Achalasia subtypes can be identified with functional luminal imaging probe (FLIP) panometry using a supervised machine learning process. *Neurogastroenterol Motil.* 2021;33(3):e13932. doi:10.1111/nmo.13932 [PubMed: 32608147]
20. Schoeman MN, Holloway RH. Stimulation and characteristics of secondary oesophageal peristalsis in normal subjects. *Gut.* 1994;35(2):152–158. [PubMed: 8307463]
21. Schoeman MN, Holloway RH. Secondary oesophageal peristalsis in patients with non-obstructive dysphagia. *Gut.* 1994;35(11):1523–1528. [PubMed: 7828966]

Key points

- Artificial intelligence (AI) could augment clinical interpretation of esophageal motility studies; this study aimed to develop an AI platform to interpret functional lumen imaging probe (FLIP) Panometry studies.
- The AI platform interpreted FLIP Panometry studies using a simple, clinically-pragmatic diagnostic scheme with accuracies of 89% compared against the impressions of experienced esophagologists.
- AI may provide useful clinical decision support for esophageal motility diagnoses from FLIP Panometry studies performed at the time of endoscopy.

**FIGURE 1.**

Flow of subjects. Flow of subjects for study inclusion and application of analysis. FLIP, functional lumen imaging probe; HRM, high-resolution manometry.

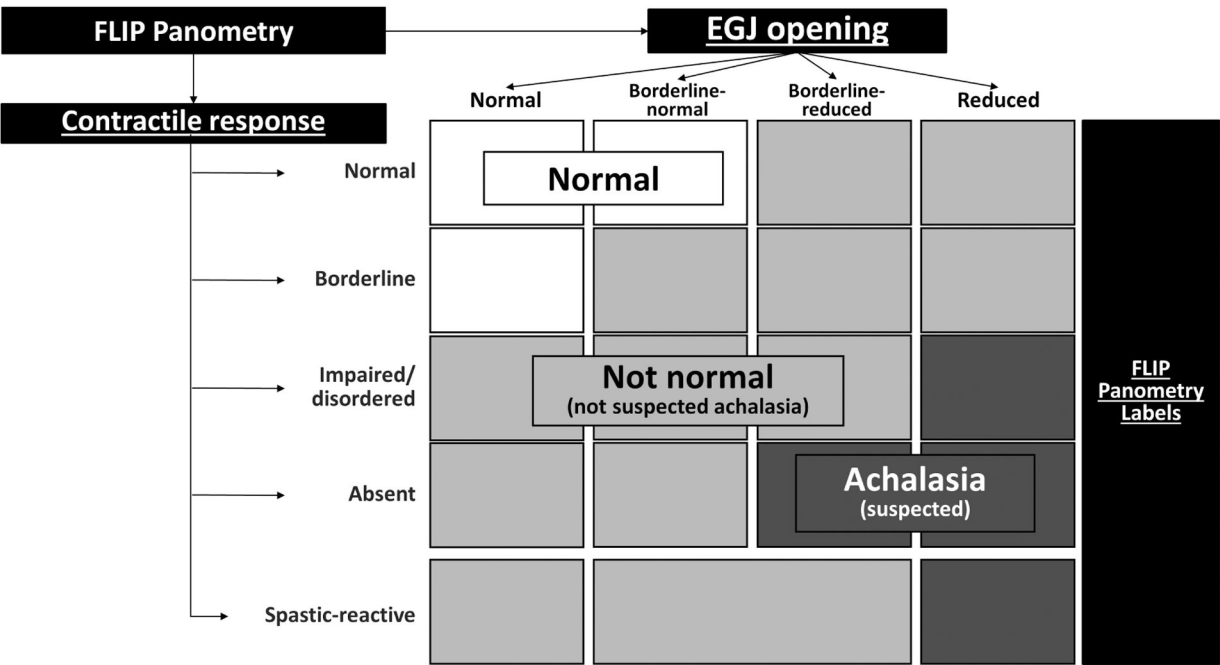


FIGURE 2. Labeling of FLIP Panometry studies for the two-stage model. The true labels for each FLIP Panometry study were assigned based on the esophagogastric junction (EGJ) opening and contractile response classifications. Figure used with permission from the Esophageal Center of Northwestern.

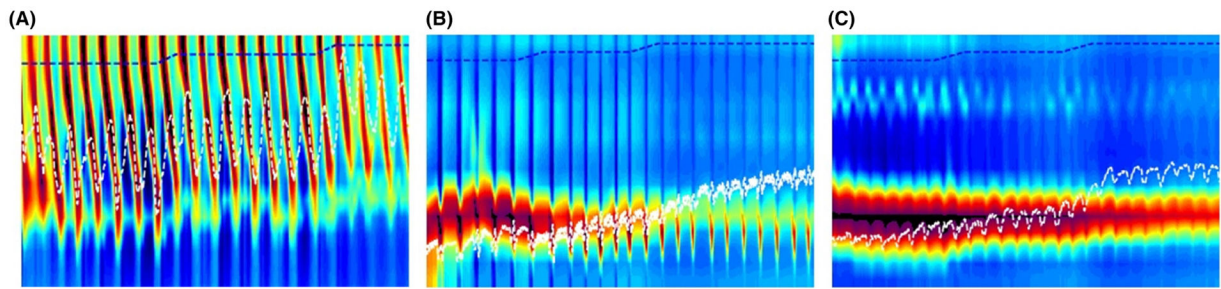


FIGURE 3.

FLIP Panometry heatmaps. The model generated heatmaps from raw FLIP data with three examples displayed (A–C). The heatmaps included the color-coded diameter by length (16 cm) by time plots with overlaid pressure (white dashed line) and FLIP fill volume (dashed blue line). The heatmaps were interpreted by the two-stage model as (A) normal, (B) not normal/not achalasia, and (C) not normal/suspected achalasia. These three patient studies were all interpreted accurately by the model. The corresponding high-resolution manometry diagnoses were (A) normal motility, (B) ineffective esophageal motility, and (C) type II achalasia. Figure used with permission from the Esophageal Center of Northwestern.

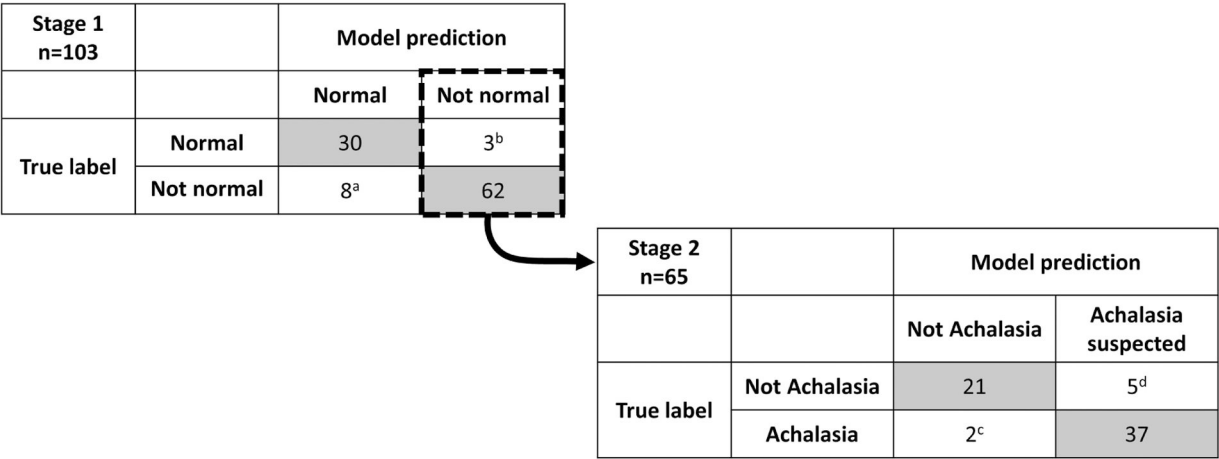


FIGURE 4. Confusion matrices for held-out test set of two-stage prediction model. Stage 1 predicted normal versus not normal, with those interpreted as “not normal” (dashed-black box) were then subjected to the stage 2 model for interpretation of suspected “achalasia” versus “not achalasia.” Gray shaded cells indicate accuracy of the model interpretation. Figure used with permission from the Esophageal Center of Northwestern. Descriptions of “inaccurate” predictions: ^aAll 8 patients had a “true” FLIP label of suspected “not achalasia” (i.e., 0/8 were suspected “achalasia”) and zero had a high-resolution manometry (HRM)/Chicago Classification v4 (CCv4.0) of achalasia. ^bAll 3/3 were subsequently predicted as “not achalasia” (i.e., 0/3 were predicted “achalasia”); all three patients had HRM with normal motility. ^cBoth patients had reduced esophagogastric junction (EGJ) opening (both with an EGJ-distensibility index of 1.9 mm²/mmHg), one with absent contractile response and the other with spastic-reactive contractile response. Both patients had type II achalasia on HRM. ^dAll 5/5 patients had an “inconclusive” FLIP motility classification with borderline EGJ opening (3/5 borderline-reduced; 2/5 borderline normal) and all 5 had abnormal contractile responses: 4/5 with impaired/disordered and 1/5 with absent. HRM/CCv4.0 diagnoses were type II achalasia in 1, EGJOO in 2 (neither completed timed barium esophagram), and normal motility in 2.

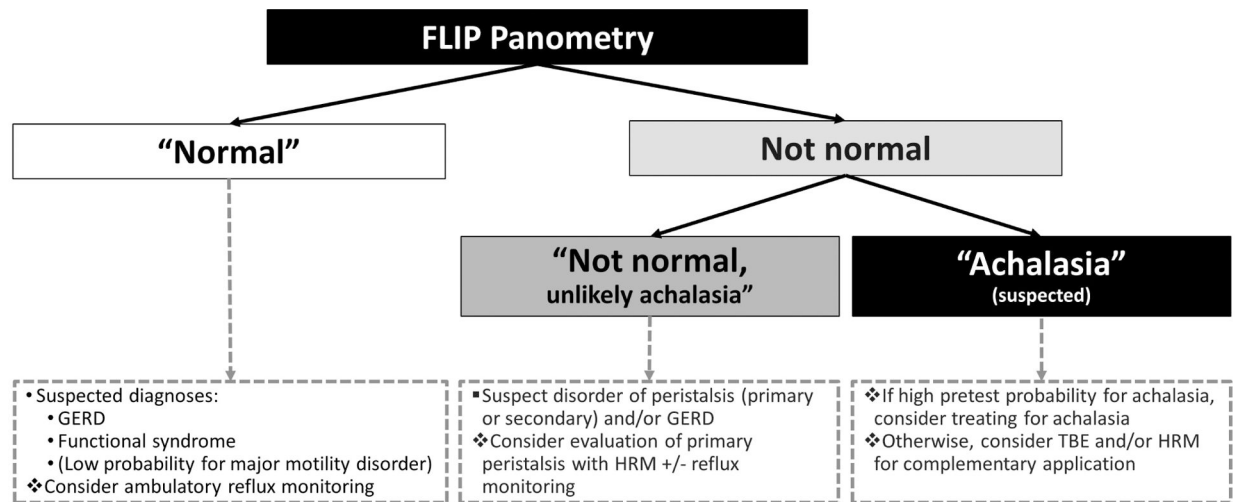


FIGURE 5.

Clinical Application of FLIP Panometry model output. The model output is intended to be applied to a pretest probability for a clinical diagnosis that is based on clinical presentation, endoscopy findings, and preexisting data (e.g., timed barium esophagram (TBE) or high-resolution manometry (HRM) if previously completed). FLIP model application is also intended for clinical scenarios in which findings associated with secondary esophageal motor abnormalities, such as large hiatal hernia, stricture, previous foregut surgery are excluded based on endoscopy and clinical history. Figure used with permission from the Esophageal Center of Northwestern.

TABLE 1

Cohort characteristics.

Variables	Patient cohort	Controls	Training cohort	Test cohort
<i>n</i> , total	678	-	610	103
Patients, <i>n</i> (%)	-	-	579 (95)	99 (96)
Controls, <i>n</i> (%)	-	35	31 (5)	4 (4)
Age, mean (SD), years	54 (17)	30 (6)	53 (17)	50 (17)
Sex, female	389 (57)	25 (71)	360 (59)	49 (52)
Indication				
Dysphagia	612 (90)	0	521 (90)	91 (92)
Reflux symptoms	40 (6)	0	35 (6)	5 (5)
Chest pain	15 (2)	0	13 (2)	2 (2)
Other	11 (2)	35 (100)	10 (2)	1 (1)
Endoscopic sedation, <i>n</i> (%)				
Conscious (midazolam/fentanyl)	544 (80)	0	495 (81)	84 (82)
Monitored anesthesia care (propofol)	134 (20)	35 (100)	115 (19)	19 (18)
FLIP Panometry (true labels)				
Two-stage classifications				
Normal	190 (28)	35	192 (32)	33 (32)
Not normal/not achalasia	182 (27)	0	151 (25)	31 (30)
Not normal/achalasia	306 (25)	0	267 (44)	39 (38)
Motility classification				
Normal	183 (27)	35 (100)	186 (31)	32 (31)
Weak	43 (6)	0	33 (5)	10 (10)
Obstruction with weak contractile response	235 (35)	0	203 (33)	32 (31)
Spastic-Reactive	77 (11)	0	67 (11)	10 (10)
Inconclusive	140 (21)	0	121 (20)	19 (18)
EGJ opening classification				
Normal	230 (34)	35 (100)	223 (37)	42 (41)
Borderline normal	80 (12)	0	69 (11)	11 (11)
Borderline-reduced	79 (12)	0	68 (11)	11 (11)
Reduced	289 (43)	0	250 (41)	39 (38)
Contractile response pattern				
Normal	105 (16)	31 (89)	113 (19)	23 (22)
Borderline	130 (19)	4 (11)	121 (20)	13 (13)
Impaired/disordered	175 (26)	0	147 (24)	28 (27)
Absent	191 (28)	0	162 (27)	29 (28)
Spastic-reactive	77 (11)	0	67 (11)	10 (10)
High-resolution manometry				
Chicago Classification v4.0				
Type I achalasia	58 (9)	0	53 (9)	5 (5)
Type II achalasia	129 (19)	0	111 (18)	18 (18)

Variables	Patient cohort	Controls	Training cohort	Test cohort
Type III achalasia	40 (6)	0	35 (6)	5 (5)
EGJOO-conclusive	18 (3)	0	15 (3)	3 (3)
EGJOO-inconclusive (inconclusive TBE)	45 (7)	0	42 (7)	3 (3)
EGJOO-inconclusive (no TBE)	76 (11)	0	61 (10)	15 (15)
Hypercontractile esophagus	15 (2)	0	12 (2)	3 (3)
Distal esophageal spasm	15 (2)	0	15 (3)	0
Absent contractility	17 (3)	0	12 (2)	5 (5)
Ineffective esophageal motility	47 (7)	3 (9)	41 (7)	9 (9)
Normal motility	218 (32)	32 (91)	213 (35)	37 (36)
Timed barium esophagram				
[<i>n</i> (%) completed TBE]	[318 (47)]	0	[274 (45)]	[44 (43)]
5min column>5cm	130 (41)	-	111 (41)	19 (43)
1min column>5cm or Tablet impaction	77 (24)		68 (25)	9 (21)
Normal	111 (35)	-	95 (35)	16 (36)

Abbreviations: EGJOO, EGJ outflow obstruction; TBE, timed barium esophagram.

Note: There were no significant differences (i.e., *p*-values >0.05) on comparisons between the training and test cohorts for any of the included variables. Values reflect *n* (%) unless otherwise specified.

Classification of esophageal motility with FLIP Panometry. These criteria were applied via manual interpretation to apply the “true” labels. The contractile response to distension was based on evaluation of the FLIP study protocol including the 50 mL, 60 mL, and 70 mL fill volumes.⁴ Esophagogastric junction (EGJ) opening applied the EGJ-distensibility index (DI) from the 60 mL FLIP fill volume and the maximum EGJ diameter from the 60 mL or 70 mL FLIP fill volume.⁴

TABLE 2

Definition	
FLIP panometry contractile response patterns	Normal contractile response
	Repetitive Antegrade Contraction (RACs), defined by the RAC Rule of 6 s:
	<ul style="list-style-type: none">• 6 consecutive antegrade contractions of• 6 cm in axial length occurring at• 6+/-3 antegrade contractions per minute regular rate
Borderline contractile response	<ul style="list-style-type: none">• Not meeting RAC Rule-of-6 s• Distinct antegrade contractions of at least 6-cm axial length present• Not spastic-reactive contractile response
	Impaired/Disordered contractile response
	<ul style="list-style-type: none">• No distinct antegrade contractions• May have sporadic or chaotic contractions not meeting antegrade contractions• Not spastic-reactive contractile response
Absent contractile response	<ul style="list-style-type: none">• No contractile activity in the esophageal body
Spastic-reactive contractile response	<ul style="list-style-type: none">• Presence of any of the following features:• Sustained occluding contractions or• Sustained lower esophageal sphincter contractions or• Repetitive retrograde contractions, defined by at least 6 consecutive retrograde contractions occurring at a rate of >9 contractions per minute
	FLIP Panometry EGJ opening classification
	Reduced EGJ opening
Borderline EGJ opening	<ul style="list-style-type: none">• EGJ-DI <2.0 mm²/mmHg AND• Maximum EGJ diameter<12 mm• EGJ-DI <2.0 mm²/mmHg OR• Maximum EGJ diameter<16 mm,• but not “reduced EGJ opening”

Definition		Further classification:
Normal EGJ opening	•	EGJ-DI 2.0 mm ² /mmHg AND
	•	Maximum EGJ diameter 16 mm
FLIP panometry motility classifications		
Normal	•	Normal EGJ opening and Normal contractile response or
	•	Normal EGJ opening and Borderline contractile response
Weak	•	Normal EGJ opening and Impaired/disordered contractile response or
	•	Normal EGJ opening and Absent contractile response
Obstruction with weak contractile response	•	Reduced EGJ opening and Impaired/disordered contractile response or
	•	Reduced EGJ opening and Absent contractile response
Spastic-reactive	•	Spastic-reactive contractile response
	•	(any EGJ opening classification)
Inconclusive	•	Reduced EGJ opening and borderline contractile response or
	•	Borderline EGJ opening and any contractile response except spastic-reactive (i.e., normal, borderline, impaired/disordered, or absent contractile response)

TABLE 3

Clinical characteristics of held-out test set based on two-stage FLIP classifier predictions

Two-stage model prediction	Normal	Not normal, Not achalasia	Not normal, Achalasia
Total, <i>n</i> (%)	38 (37)	23 (22)	42 (41)
Patients *, <i>n</i> (%)	34 (34)	23 (23)	42 (42)
Controls *, <i>n</i> (%)	4 (100)	0	0
Indication (<i>n</i> , % patients)			
Dysphagia *	29 (85)	21 (91)	41 (97) ^a
Reflux symptoms	4 (12)	1 (4)	0
Chest pain	1 (3)	0	1 (2)
Other	0	1 (4)	0
FLIP Panometry (true labels)			
Motility classification *			
Normal	29 (73)	3 (13) ^a	0 ^{a,b}
Weak	6 (16)	4 (17)	0 ^{a,b}
Obstruction with weak contractile response	0	1 (4)	31 (74) ^{a,b}
Spastic-Reactive	0	6 (26) ^a	4 (10)
Inconclusive	3 (8)	9 (39) ^a	7 (17)
EGJ opening classification *			
Normal	35 (92)	7 (30) ^a	0 ^{a,b}
Borderline normal	3 (8)	6 (26) ^a	2 (5) ^a
Borderline-reduced	0	6 (26) ^a	5 (12)
Reduced	0	4 (17) ^a	35 (83) ^{a,b}
Contractile response pattern *			
Normal	23 (61)	0 ^a	0 ^a
Borderline	7 (18)	6 (26)	0 ^{a,b}
Impaired/disordered	7 (18)	7 (30)	14 (33)
Absent	1 (3)	4 (17)	24 (57) ^{a,b}
Spastic-reactive	0	6 (26) ^a	4 (10)
High-resolution manometry			
Chicago Classification v4.0 *			
Type I achalasia	0	0	5 (12) ^{a,b}
Type II achalasia	0	2 (9)	16 (38) ^{a,b}
Type III achalasia	0	0	5 (12)
EGJOO-conclusive	0	0	3 (7)
EGJOO-inconclusive (inconclusive TBE)	1 (3)	2 (9)	0
EGJOO-inconclusive (no TBE)	2 (5)	6 (26)	7 (17)
Hypercontractile esophagus	2 (5)	1 (4)	0

Two-stage model prediction	Normal	Not normal, Not achalasia	Not normal, Achalasia
Distal esophageal spasm	0	0	0
Absent contractility	4 (11)	1 (4)	0
Ineffective esophageal motility	3 (8)	4 (17)	2 (5)
Normal motility	26 (68)	7 (30) ^a	4 (10) ^a
Timed barium esophagram			
[n (%) completed TBE]	13 (34)	8 (35)	23 (55)
5 min column >5 cm *	1 (8)	1 (13)	17 (74) ^a
1 min column >5 cm or Tablet impaction	4 (31)	2 (25)	3 (13)
Normal *	8 (62)	5 (63)	3 (13) ^a

Note: Values reflect *n* (%) of predicted label unless otherwise specified.

On post-hoc, pairwise comparisons after Bonferroni correction:

^a *p* < 0.05 on comparison with “Normal;”

^b *p* < 0.05 on comparison with “Not normal; Not achalasia.”

Abbreviations: EGJOO, EGJ outflow obstruction; TBE, timed barium esophagram.

* *p* < 0.05 on comparison across three groups.

TABLE 4

Cases with “inaccurate” two-stage model predictions.

Model prediction	“True” label	FLIP motility classification	EGJ opening classification	Contractile response pattern	EGJ-DI (mm ² /mmHg)	Maximum EGJ diameter (mm)	HRM/CCv4.0 diagnosis
Normal	Not achalasia	Weak	Normal	Impaired/disordered	5.4	19.7	Normal
Normal	Not achalasia	Weak	Normal	Impaired/disordered	5.5	16.4	Normal
Normal	Not achalasia	Weak	Normal	Impaired/disordered	2.9	18.5	Normal
Normal	Not achalasia	Weak	Normal	Impaired/disordered	2.6	17.2	EGJOO—inconclusive ^a
Normal	Not achalasia	Weak	Normal	Impaired/disordered	7.6	19.1	Absent contractility
Normal	Not achalasia	Weak	Normal	Absent	5.9	17.5	Absent contractility
Normal	Not achalasia	Inconclusive	Borderline normal	Impaired/disordered	4.3	15.2	Normal
Normal	Not achalasia	Inconclusive	Borderline normal	Impaired/disordered	4.0	14.7	Absent
Not achalasia	Normal	Normal	Normal	Borderline	2.2	18.5	Normal
Not achalasia	Normal	Normal	Normal	Borderline	3.9	21.5	Normal
Not achalasia	Normal	Normal	Normal	Borderline	3.6	19.2	Normal
Not achalasia	Achalasia	Spastic-reactive	Reduced	Spastic-reactive	1.9	9.8	Type II achalasia
Not achalasia	Achalasia	Obstruction w/weak contractile response	Reduced	Absent	1.9	11.1	Type II achalasia
Achalasia	Not achalasia	Inconclusive	Borderline-reduced	Impaired/disordered	0.8	12.5	Type II achalasia
Achalasia	Not achalasia	Inconclusive	Borderline-reduced	Impaired/disordered	2.2	11.6	Normal
Achalasia	Not achalasia	Inconclusive	Borderline normal	Absent	2.7	15.1	Normal
Achalasia	Not achalasia	Inconclusive	Borderline normal	Impaired/disordered	1.2	14.9	EGJOO—inconclusive ^b
Achalasia	Not achalasia	Inconclusive	Borderline-reduced	Impaired/disordered	1.7	13.8	EGJOO—inconclusive ^b

Note: Each row represents one patient/FLIP study with an inaccurate model prediction.

Abbreviations: CCv4.0, Chicago Classification version 4.0; EGJ, esophagogastric junction; DI, distensibility index; HRM, high-resolution manometry.

^aTimed barium esophagram (TBE) was normal.

^bTBE was not completed.